

On the Myopic Policy for a Class of Restless Bandit Problems with Applications in Dynamic Multichannel Access

Keqin Liu and Qing Zhao

Abstract

We consider a class of restless multi-armed bandit problems that arises in multi-channel opportunistic communications, where channels are modeled as independent and stochastically identical Gilbert-Elliot channels and channel state observations are subject to errors. We show that the myopic channel selection policy has a semi-universal structure that obviates the need to know the Markovian transition probabilities of the channel states. Based on this semi-universal structure, we establish closed-form lower and upper bounds on the maximum throughput (*i.e.*, average reward) achieved by the myopic policy. Furthermore, we characterize the approximation factor of the myopic policy by considering a genie-aided system.

Index Terms

Dynamic multi-channel access, restless multi-armed bandit, myopic policy

I. INTRODUCTION

A. Dynamic Multichannel Access

We consider the following stochastic optimization problem that arises in multichannel opportunistic communications. Assume that there are N independent and stochastically identical Gilbert-Elliot channels [1]. As illustrated in Fig. 1, the state of a channel — “good” or “bad” — indicates the desirability of accessing this channel and determines the resulting reward. The transitions between these two states follow a discrete-time Markov chain with transition

This work was supported by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grants ECS-0622200 and CCF-0830685.

Keqin Liu and Qing Zhao are with the Department of Electrical and Computer Engineering, University of California, Davis, CA, 95616, USA {kqliu, qzhao}@ucdavis.edu

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE MAR 2009		2. REPORT TYPE		3. DATES COVERED 00-00-2009 to 00-00-2009	
4. TITLE AND SUBTITLE On the Myopic Policy for a Class of Restless Bandit Problems with Applications in Dynamic Multichannel Access				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of California, Department of Electrical and Computer Engineering, Davis, CA, 95616				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT We consider a class of restless multi-armed bandit problems that arises in multi-channel opportunistic communications, where channels are modeled as independent and stochastically identical Gilbert- Elliot channels and channel state observations are subject to errors. We show that the myopic channel selection policy has a semi-universal structure that obviates the need to know the Markovian transition probabilities of the channel states. Based on this semi-universal structure, we establish closed-form lower and upper bounds on the maximum throughput (i.e., average reward) achieved by the myopic policy. Furthermore, we characterize the approximation factor of the myopic policy by considering a genie-aided system.					
15. SUBJECT TERMS Dynamic multi-channel access, restless multi-armed bandit, myopic policy					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 27	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

probabilities $\{p_{ij}\}_{i,j \in \{0,1\}}$. This channel model has been commonly used to abstract physical channels with memory. Consider, for example, the emerging application of cognitive radios for opportunistic spectrum access where secondary users search in the spectrum for idle channels temporarily unused by primary users [2]. For this application, the good state represents an idle channel while the bad state an occupied channel. When the primary network employs load balancing across channels, the occupancy processes of all channels can be considered stochastically identical.

In each time slot, a user chooses M out of the N channels to sense and subsequently access channels sensed to be in the good states. Sensing is subject to errors: a good channel may be sensed as bad and *vice versa*. Accessing a good channel results in a unit reward, and no access or accessing a bad channel leads to zero reward. The design objective is the optimal sensing policy for dynamic channel selection in order to maximize the expected long-term reward.

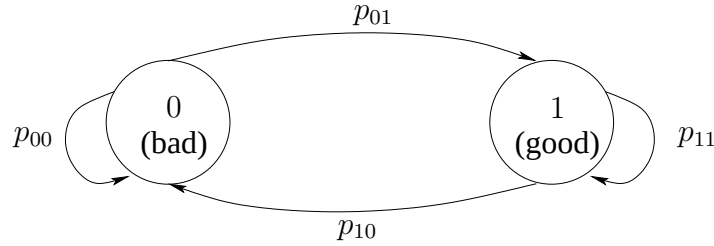


Fig. 1. The Gilbert-Elliott channel model.

B. Restless Multi-armed Bandit and Myopic Policy

This problem can be formulated as a partially observable Markov decision process (POMDP) for generally correlated channels [3], or a restless multi-armed bandit process (RMBP) for independent channels considered here. The maximum throughput of the multi-channel opportunistic system is essentially the long-term expected maximum average reward, or the time-normalized value function, of an RMBP. Unfortunately, obtaining optimal solutions to a general restless bandit process is PSPACE-hard [4], and analytical characterization of the performance of the optimal policy is often intractable.

We thus focus on the low-complexity myopic policy which has been shown to be optimal for this class of restless bandit problems under certain conditions (see Sec. I-C). Specifically, we

establish a semi-universal structure of the myopic policy and characterize its performance and approximation factor as detailed below.

1) *Structure of the Myopic Policy:* We first show that the myopic policy has a semi-universal structure under the condition that the probability of false alarm of the channel state detector is below a certain value. This structure reveals that the myopic policy does not require the knowledge of the transition probabilities of the Markovian channel model except the order of p_{11} and p_{01} .

2) *Performance of the Myopic Policy:* Based on the semi-universal structure of the myopic policy, we develop closed-form lower and upper bounds on the steady-state throughput under the myopic policy that monotonically tighten as the number N of channels increases. When each channel is positively correlated ($p_{11} \geq p_{01}$), we further obtain the limiting performance of the myopic policy as N approaches to infinity.

3) *Approximation Factor of the Myopic Policy:* By considering a genie-aided system, we develop an upper bound on the optimal performance, which provides a performance benchmark for the myopic policy. This result, coupled with the lower bound on the performance of the myopic policy, leads to an analytical characterization of the approximation factor of the myopic policy. Specifically, we show that the myopic policy achieves at least $\frac{M}{N}$ of the optimal performance when channels are positively correlated, and $\max\{\frac{1}{2}, \frac{M}{N}\}$ of the optimal performance when channels are negatively correlated ($p_{11} < p_{01}$).

C. Related Work

1) *Perfect State Observation:* Under the assumption of single-channel perfect sensing, the semi-universal structure of the myopic policy has been established for all N , and the optimality of the myopic policy proved for $N = 2$ and conjectured for $N > 2$ in [5]. Furthermore, closed-form bounds on the throughput under the myopic policy have been established. A recent follow-up work [6] has extended the optimality of the myopic policy to all N under the condition of $p_{11} \geq p_{01}$.

For independent and non-identical channels under multi-channel perfect sensing, Whittle's index policy under both discounted and average reward criteria has been established in [7]. An efficiently computable upper bound on the optimal performance has been established based on Whittle's relaxation. Numerical results have illustrated the strong performance of Whittle's

index policy. For independent and identical channels, Whittle's index policy has been shown to be equivalent to the myopic policy. The structure of the myopic policy under multi-channel sensing has been established, and the myopic policy has been shown to be optimal when $M = N - 1$. Furthermore, an approximation factor of the myopic policy has been developed for general M and N . Interestingly, the approximation factor we establish in this paper coincides with the one obtained in [7].

2) *Imperfect State Observation*: Under imperfect sensing, the design of multi-channel opportunistic access was addressed in [8] under a general correlated channel model. This problem requires the joint design of the channel state detector, the access policy and the sensing policy. A separation principle has been established which decouples the design of channel state detector and access policies from that of channel sensing policy. The channel sensing policy then falls into an unconstrained POMDP problem. In [9], the structure and optimality of the myopic sensing policy has been established under certain conditions for independent and identical channels. Specifically under single-channel sensing, a simple and robust round-robin structure of the myopic policy has been established when the false alarm probability of the channel state detector is below a certain value. Based on this structure, the myopic policy has been shown to be optimal for $N = 2$. In this paper, we extend the structure of the myopic policy to multi-channel sensing scenarios.

II. PROBLEM FORMULATION

A. System Model

Let $\mathcal{S}(t) \triangleq [S_1(t), \dots, S_N(t)]$ denote the channel states, where $S_n(t) \in \{0 \text{ (bad), } 1 \text{ (good)}\}$ is the state of channel n in slot t . At the beginning of each slot, the user first decides which M channels to sense for potential access. Once a channel (say channel n) is chosen, the user detects the channel state, which can be considered as a binary hypothesis test¹:

$$\mathcal{H}_0 : S_n(t) = 1 \text{ (good)} \text{ vs. } \mathcal{H}_1 : S_n(t) = 0 \text{ (bad)}.$$

The performance of channel state detection is characterized by the ROC which relates the probability of false alarm ϵ and the probability of miss detection δ :

$$\epsilon \triangleq \Pr\{\text{decide } \mathcal{H}_1 | \mathcal{H}_0 \text{ is true}\}, \quad \delta \triangleq \Pr\{\text{decide } \mathcal{H}_0 | \mathcal{H}_1 \text{ is true}\}.$$

¹We consider here the nontrivial cases with p_{01} and p_{11} in the open interval of $(0,1)$. When they take the special value of 0 or 1, channel state detection can be simplified. Extensions to such special cases are straightforward.

Based on the imperfect detection outcome in slot t , the user chooses an access action $\Phi_n(t) \in \{0 \text{ no access, } 1 \text{ access}\}$ that determines whether to access channel n for transmission. We note that the design should be subject to a constraint on the probability of accessing a bad channel, which may cause interference or waste energy. Specifically, the probability of collision $\mathcal{P}_n(t)$ perceived by the primary network in any channel and slot is capped below a predetermined threshold ζ , *i.e.*,

$$\mathcal{P}_n(t) \triangleq \Pr(\Phi_n(t) = 1 | S_n(t) = 0) \leq \zeta, \quad \forall n, t.$$

This constrained stochastic optimization problem requires the joint design of the channel state detector (*i.e.*, how to choose the detection thresholds to trade off false alarms with miss detections), the access policy that decides the transmission probabilities based on imperfect detection outcomes, and the sensing policy for channel selection. This problem is formulated as a constrained POMDP in [8] for generally correlated channels. A separation principle has been established that the optimal detector is the Neyman-Pearson detector with the probability δ of miss detection given by the maximum allowable probability ζ of collision, and the optimal access policy is to simply trust the detection outcomes: transmit over a channel if and only if it is detected as good. Thus, the user can obtain a unit reward on a chosen channel if and only if it is in good state and detected correctly (*i.e.*, no false alarm). The optimal sensing policy can then be designed using the optimal detector and the optimal access policy without the constraint on accessing a bad channel, which becomes an unconstrained POMDP addressed here. The objective is to maximize the average reward (throughput) over a horizon of T slots by choosing judiciously a sensing policy that governs channel selection in each slot.

Since failed transmission may occur, acknowledgements (ACKs) are necessary to ensure guaranteed delivery. Specifically, when the receiver successfully receives a packet from a channel, it sends an acknowledgement to the transmitter over the same channel at the end of the slot. Otherwise, the receiver does nothing, *i.e.*, a NAK is defined as the absence of an ACK, which occurs when the transmitter did not transmit over this channel or transmitted but the channel is in bad state. We assume that acknowledgements are received without error since acknowledgements are always transmitted over good/idle channels.

B. Restless Multi-Armed Bandit Formulation

Due to limited and imperfect sensing, the system state $[S_1(t), \dots, S_N(t)] \in \{0, 1\}^N$ in slot t is not fully observable to the user. It can, however, infer the state from its decision and observation history. It has been shown that a sufficient statistic of the system for optimal decision making is given by the conditional probability that each channel is in state 1 given all past decisions and observations [10]. Referred to as the belief vector, this sufficient statistic is denoted by $\Omega(t) \triangleq [\omega_1(t), \dots, \omega_N(t)]$, where $\omega_i(t)$ is the conditional probability that $S_i(t) = 1$. In order to ensure that the user and its intended receiver tune to the same channels in each slot, channel selections should be based on common observations: the acknowledgements $\mathcal{K}(t) \in \{0 \text{ (NAK)}, 1 \text{ (ACK)}\}^M$ in each slot rather than the detection outcomes at the transmitter. Let $I(t)$ denote the sensing action that consists of M channels to sense in slot t . Given the sensing action $I(t)$ and the observations $\{K_i(t) \in \{0, 1\} : i \in I(t)\}$ in slot t , the belief vector for slot $t + 1$ can be obtained via the Bayes rule.

$$\omega_i(t+1) = \begin{cases} p_{11}, & i \in I(t), K_i(t) = 1 \\ \Gamma(\frac{\epsilon\omega_i(t)}{\epsilon\omega_i(t)+1-\omega_i(t)}), & i \in I(t), K_i(t) = 0 \\ \Gamma(\omega_i(t)), & i \notin I(t) \end{cases} \quad (1)$$

where the operator $\Gamma(\cdot)$ is defined as

$$\Gamma(x) \triangleq xp_{11} + (1-x)p_{01}.$$

A sensing policy π specifies a sequence of functions $\pi = [\pi_1, \pi_2, \dots, \pi_T]$ where π_t maps a belief vector $\Omega(t)$ to a sensing action $I(t)$ for slot t . Multi-channel opportunistic access can thus be formulated as the following stochastic optimization problem.

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=1}^T R(\pi_t(\Omega(t))) | \Omega(1) \right],$$

where $R(\pi_t(\Omega(t)))$ is the reward obtained when the belief is $\Omega(t)$ and channels $\pi_t(\Omega(t))$ are selected, and $\Omega(1)$ is the initial belief vector. This problem falls into the model of an RMBP by treating the belief value of each channel as the state of each arm of a bandit. If no information on the initial system state is available, each entry of $\Omega(1)$ can be set to the stationary distribution ω_o of the underlying Markov chain:

$$\omega_o = \frac{p_{01}}{p_{01} + p_{10}}. \quad (2)$$

Let $V_t(\Omega)$ be the value function, which represents the maximum expected total reward that can be obtained starting from slot t given the current belief vector Ω . Given that the user takes action I and observes $\mathcal{K} = \{K_i\}_{i \in I}$, the expected reward that can be accumulated starting from slot t consists of two parts: the expected immediate reward $\sum_{i \in I} \omega_i (1 - \epsilon)$ and the maximum expected future reward $V_{t+1}(\mathcal{T}(\Omega|I, \mathcal{K}))$, where $\mathcal{T}(\Omega|I, \mathcal{K})$ denotes the updated belief vector for slot $t + 1$ after incorporating action I and observations $\mathcal{K}$ as given in (1). Averaging over all possible observations $\mathcal{K}$ and maximizing over all actions I , we arrive at the following optimality equation.

$$\begin{aligned} V_T(\Omega(T)) &= \max_I \sum_{i \in I} \omega_i (1 - \epsilon), \\ V_t(\Omega(t)) &= \max_I (\sum_{i \in I} \omega_i (1 - \epsilon) + \mathbb{E}[V_{t+1}(\mathcal{T}(\Omega(t)|I, \mathcal{K}))]). \end{aligned}$$

In theory, the optimal policy π^* and its performance $V_1(\Omega(1))$ can be obtained by solving the above dynamic programming. Unfortunately, due to the impact of the current action on the future reward and the uncountable space of the belief vector, obtaining the optimal solution using directly the above recursive equations is computationally prohibitive. Even when approximate numerical solutions can be obtained, they do not provide insight into system design or analytical characterizations of the optimal performance $V_1(\Omega(1))$.

III. STRUCTURE AND PERFORMANCE OF THE MYOPIC POLICY

A. Myopic Policy

A myopic policy ignores the impact of the current action on the future reward, focusing solely on maximizing the expected immediate reward $\mathbb{E}[R(I(t))]$. Myopic policies are thus stationary. The myopic action $\hat{I}$ under belief state $\Omega = [\omega_1, \dots, \omega_N]$ is simply given by

$$\hat{I}(\Omega) = \arg \max_I \sum_{i \in I} \omega_i. \quad (3)$$

In general, obtaining the myopic action in each slot requires the recursive update of the belief vector Ω as given in (1), which requires the knowledge of the transition probabilities $\{p_{ij}\}$. Interestingly, it has been shown in [9] under single-channel sensing ($M = 1$) that the myopic policy has a simple structure that does not need the update of the belief vector or the precise knowledge of the transition probabilities if the probability of false alarm is below a certain value.

Surprisingly, the myopic policy with such a simple and robust structure achieves the optimal performance for $N = 2$ [9]. Under multi-channel sensing ($M \geq 1$), extensive simulations have shown that the myopic policy achieves the optimal performance. We thus conjecture that the optimality of the myopic policy holds for general M and N . In the next section, we show that the structure of the myopic policy can be directly generalized to multi-channel sensing scenarios. Based on this structure, we characterize the performance of the myopic policy.

B. Structure

We first present the following assumptions.

A1: The initial belief values are bounded between p_{01} and p_{11} .

A2:

$$\epsilon \leq \frac{\min\{p_{01}, p_{11}\}(1 - \max\{p_{01}, p_{11}\})}{\max\{p_{01}, p_{11}\}(1 - \min\{p_{01}, p_{11}\})}.$$

Assumption A1 will only be used in Theorem 1 which describes the structure of the myopic policy. We note that the structure can be directly extended if assumption A1 does not hold. We assume A1 in Theorem 1 for the easy of presentation.

For Assumption A2, the allowed probability of miss detection δ plays a major role since ϵ can be reduced to an arbitrarily small value at the price of increased δ . However, both ϵ and δ can be improved by increasing the sensing/detection time (*i.e.*, taking more measurements). The caveat is the reduced transmission time for a given slot length. This interesting tradeoff between the complexity of the detector at the physical layer and the transmission strategy at the Medium Access Control (MAC) layer of a communication network can be complex and is beyond the scope of this paper.

The implementation of the myopic policy can be described with a queue structure. Specifically, all N channels are ordered in a queue, and in each slot, those M channels at the head of the queue are sensed.

Theorem 1: The Semi-Universal structure of the myopic policy

The initial channel ordering $\mathbf{Q}(1)$ is determined by the initial belief vector as given below.

$$\omega_{n_1}(1) \geq \cdots \geq \omega_{n_N}(1) \implies \mathbf{Q}(1) = (n_1, \cdots, n_N).$$

Under assumption A1 and A2, channels are reordered at the end of each slot according to the following simple rules. When $p_{11} \geq p_{01}$, the channels observed with ACK will stay at the head of the queue, and the channels observed with NAK will be moved to the end of the queue while keeping their order unchanged. When $p_{11} < p_{01}$, the channels observed with NAK will stay at the head of the queue while reversing their order, and the channels observed with ACK will be moved to the end of the queue. The order of the unobserved channels are also reversed.

Proof: Let $\mathbf{Q}(t) = (n_1, n_2, \dots, n_N)$ ($n_i \in \{1, 2, \dots, N\} \forall i$) be the queueing order of channels in slot t . We need to show that

$$\omega_{n_1}(t) \geq \dots \geq \omega_{n_N}(t). \quad (4)$$

We first present the following properties of the operator $\Gamma(x)$ defined in (1).

- P1. $\Gamma(x)$ is an increasing function for $p_{11} \geq p_{01}$ and a decreasing function for $p_{11} < p_{01}$.
- P2. $\forall 0 \leq x \leq 1$, $p_{01} \leq \Gamma(x) \leq p_{11}$ for $p_{11} \geq p_{01}$ and $p_{11} \leq \Gamma(x) \leq p_{01}$ for $p_{11} < p_{01}$.
- P3. For $p_{11} \geq p_{01}$ and $\epsilon \leq \frac{p_{10}p_{01}}{p_{11}p_{00}}$, we have $\Gamma(\frac{\epsilon\omega}{\epsilon\omega+(1-\omega)}) \leq \Gamma(\omega') \forall p_{01} \leq \omega, \omega' \leq p_{11}$; for $p_{11} < p_{01}$ and $\epsilon \leq \frac{p_{00}p_{11}}{p_{01}p_{10}}$, we have $\Gamma(\frac{\epsilon\omega}{\epsilon\omega+(1-\omega)}) \geq \Gamma(\omega') \forall p_{11} \leq \omega, \omega' \leq p_{01}$.

P1 and P2 follow directly from the definition of $\Gamma(x)$. To show P3 for $p_{11} \geq p_{01}$, it suffices to show $\frac{\epsilon\omega}{\epsilon\omega+(1-\omega)} \leq p_{01}$ due to the monotonically increasing property of $\Gamma(x)$ and the bound on ω' . Noticing that $\frac{\epsilon\omega}{\epsilon\omega+(1-\omega)}$ is an increasing function of both ω and ϵ , we arrive at P3 by using the upper bounds on ω and ϵ . Similarly, we can show P3 for $p_{11} < p_{01}$.

We now prove (4) by induction. For $t = 1$, (4) holds by the definition of $\mathbf{Q}(1)$. Assume that (4) is true for slot t . We show that it is also true for slot $t + 1$.

Consider first $p_{11} \geq p_{01}$. For an $1 \leq i \leq M$ with $K_{n_i} = 1$, $\omega_{n_i}(t + 1) = p_{11}$ which achieves the upper bound of the belief values (See P2). For an $1 \leq j \leq M$ with $K_{n_j} = 0$, $\omega_{n_j}(t + 1)$ is upper bounded by those of unobserved channels due to P3. Among those channels observed 0, the order of their believes remains unchanged in slot $t + 1$ due to P1. Similarly, the order of the belief values of the unobserved channels also remains unchanged in slot $t + 1$.

For $p_{11} < p_{01}$, the belief values of channels observed 1 will achieve the lower bound p_{11} of the belief values (See P2). For an $1 \leq j \leq M$ with $K_{n_j} = 0$, $\omega_{n_j}(t + 1)$ is lower bounded by those of unobserved channels due to P3. Among those channels observed 0, the order of their believes will be reversed in slot $t + 1$ due to P1. Similarly, the order of the belief values of the unobserved channels will also be reversed in slot $t + 1$.

We thus proved (4) for all $t \geq 1$ under the structure of the myopic policy. ■

Based on this structure, the myopic policy can be implemented without knowing the channel transition probabilities except the order of p_{11} and p_{01} . As a result, the myopic policy is robust against model mismatch and automatically tracks variations in the channel model provided that the order of p_{11} and p_{01} remains unchanged. Following the *belief-independence* property of this simple structure, we present the following corollary which allows us to work with a Markov reward process with a finite state space instead of one with an uncountable state space (*i.e.*, belief vectors) as we encounter in a general POMDP.

Corollary 1: Let $\mathbf{Q}(t) = (n_1, n_2, \dots, n_N)$ ($n_i \in \{1, 2, \dots, N\} \forall i$) be the queueing order of channels in slot t , where myopic action $\hat{I}(t) = \{n_i\}_{i=1}^M$. Define $\vec{\mathbf{S}}(t) \triangleq [S_{n_1}(t), S_{n_2}(t), \dots, S_{n_N}(t)]$ and $\vec{\mathbf{E}}(t) \triangleq \{e_1(t), e_2(t), \dots, e_M(t)\}$, where $\{e_i(t)\}_{1 \leq i \leq M, t \geq 1}$ are i.i.d. binary random variables taking value 0 with probability ϵ and 1 with probability $1 - \epsilon$. Under assumption A2, the augmented Markov process $\vec{\mathbf{G}}(t) \triangleq [\vec{\mathbf{S}}(t), \vec{\mathbf{E}}(t)]$ form a 2^{N+M} -state Markov chain, and the performance of the myopic policy is determined by the Markov reward process $(\vec{\mathbf{G}}(t), R(t))$ with $R(t) = \sum_{i=1}^M S_{n_i}(t)e_i(t)$.

Proof: $\vec{\mathbf{G}}(t)$ specifies the states of all channels, the queueing order of channels under the myopic policy, and the observations obtained in slot t . Specifically, the observation (0 (NAK), 1 (ACK)) on channel n_i ($1 \leq i \leq M$) in slot t is given by $S_{n_i}(t)e_i(t)$. Based on the structure of the myopic policy, $\vec{\mathbf{G}}(t)$ determines the probability distribution of $\vec{\mathbf{G}}(t+1)$, *i.e.*, $\vec{\mathbf{G}}(t)$ is a Markov chain. Furthermore, the reward $R(t)$ in slot t is given by the number of channels observed with ACK. ■

Theorem 1 and Corollary 1 provides foundations in analyzing the performance of the myopic policy.

C. Performance

In this section, we analyze the performance of the myopic policy. Under the optimality conjecture (see Sec. III-A), the throughput achieved by the myopic policy defines the performance limit of a multi-channel opportunistic communications system. In particular, we are interested in the relationship between the throughput achieved by the myopic policy and the number N of channels.

1) *Uniqueness of Steady-State Performance and Its Numerical Evaluation:* We first establish the existence and uniqueness of the system steady-state performance under the myopic policy. The steady-state throughput under the myopic policy is given by

$$U(\Omega(1)) \triangleq \lim_{T \rightarrow \infty} \frac{\hat{V}_{1:T}(\Omega(1))}{T}, \quad (5)$$

where $\hat{V}_{1:T}(\Omega(1))$ is the expected total reward obtained in T slots under the myopic policy when the initial belief is $\Omega(1)$. From Corollary 1, $U(\Omega(1))$ is determined by the Markov reward process $\{\vec{G}(t), R(t)\}$. It is easy to see that the 2^{N+M} -state Markov chain $\{\vec{G}(t)\}$ is irreducible and aperiodic, thus has a limiting distribution. As a consequence, the limit in (5) exists, and the steady-state throughput U is independent of the initial belief value $\Omega(1)$.

Corollary 1 also provides a numerical approach to evaluating U by calculating the limiting (stationary) distribution of $\vec{G}(t)$ whose transition probabilities can be directly obtained from the transition probabilities of the channel states. This numerical approach, however, does not provide an analytical characterization of the throughput U in terms of the number N of channels and the transition probabilities $\{p_{i,j}\}$. In the next section, we obtain analytical expressions of U and its scaling behavior with respect to N based on a stochastic dominance argument.

2) *Analytical Characterization of Throughput:* From the structure of the myopic policy, the throughput is determined by how often the user switches channels. When $p_{11} \geq p_{01}$, the event of a channel switching is equivalent to a slot *without* reward. The opposite holds when $p_{11} < p_{01}$: a channel switching corresponds to a slot *with* reward. For both cases, we note that the user may switch to the same channel when a channel switch is needed.

We thus introduce the concept of *transmission period (TP)*, which is the time period starting from the slot the user switches to a channel and ending at the slot that the next switch on this channel is needed (see Fig. 2 for an example under single-channel sensing). Note that the user may switch to the same channel. We count the transmission periods in the order of its starting point. Let L_k denote the length of the k th TP. We then have a discrete-time random process $\{L_k\}_{k=1}^{\infty}$ with a state space of positive integers.

Lemma 1:

$$U = \begin{cases} M(1 - 1/\bar{L}), & p_{11} \geq p_{01} \\ M/\bar{L}, & p_{11} < p_{01} \end{cases}. \quad (6)$$

where $\bar{L} = \lim_{K \rightarrow \infty} \frac{\sum_{k=1}^K L_k}{K}$ denotes the average length of a TP.

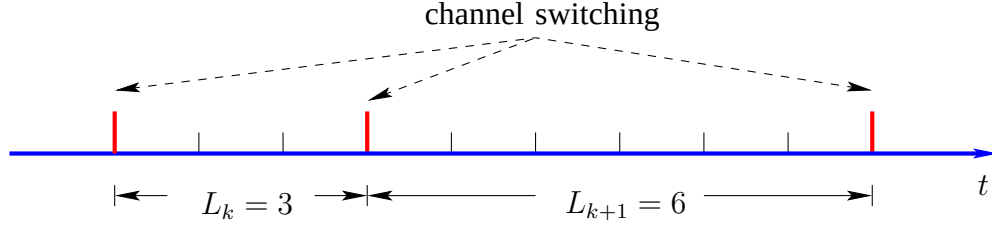


Fig. 2. The transmission period structure.

Proof: Consider first $p_{11} \geq p_{01}$. Let k_T denote the number of channel switches during a finite horizon of length T . Since a channel switch represents a loss of a unit of reward, the throughput U_T during the finite horizon is given below.

$$U_T = \frac{MT - k_T}{T}. \quad (7)$$

Let j_T denote the number of TPs during the finite horizon. We have $j_T = M + k_T$ since a channel switch initializes a new TP. It is easy to see that $MT \leq \sum_{i=1}^{M+k_T} L_k \leq MT + \sum_{k+1}^{M+k_T} L_k$. Note that the length of a TP is finite almost surely. We thus have

$$\lim_{T \rightarrow \infty} \frac{MT}{k_T} = \lim_{T \rightarrow \infty} \frac{\sum_{i=1}^{M+k_T} L_k}{M + k_T} = \bar{L}. \quad (a.s.) \quad (8)$$

From (7) and (8), we have

$$U = \lim_{T \rightarrow \infty} U_T = M \lim_{T \rightarrow \infty} \left(1 - \frac{k}{MT}\right) = M(1 - 1/\bar{L}). \quad (a.s.) \quad (9)$$

The case for $p_{11} < p_{01}$ can be similarly obtained by observing that a channel switch represents a gain of one unit reward. ■

Based on Lemma 1, throughput analysis is reduced to analyzing the average TP length $\bar{L}$. We note that the distribution of L_k is determined by the belief value in the first slot of the k -th TP. Under single-channel sensing ($M = 1$), the approach is to construct first-order Markov chains that stochastically dominate or are dominated by $\{L_k\}_{k=1}^{\infty}$. The stationary distributions of these first-order Markov chains, which can be obtained in closed-form, lead to lower and upper bounds on U according to (6). Specifically, for $p_{11} \geq p_{01}$, a lower bound on U is obtained by constructing a first-order Markov chain whose stationary distribution is stochastically dominated by the stationary distribution of $\{L_k\}_{k=1}^{\infty}$. An upper bound on U is given by a first-order Markov chain whose stationary distribution stochastically dominates the stationary distribution of $\{L_k\}_{k=1}^{\infty}$. Similarly, bounds on U can be obtained for $p_{11} < p_{01}$.

Theorem 2: Define functions

$$f(x) \triangleq \frac{\omega_o - x}{1 - x(1 - \epsilon)\left(1 - \frac{(p_{11} - p_{01})(1 - p_{11}(1 - \epsilon))}{1 - (p_{11} - p_{01})p_{11}(1 - \epsilon)}\right)},$$

$$h(x, y, z, a, b) \triangleq \frac{1 - \omega_o(1 - \epsilon) + a}{1 - a\left(\frac{(y(p_{11} - p_{01})^2 + (p_{11} - p_{01})^{b+1})z}{1 - ((p_{11} - p_{01})y)^2} - x\right)},$$

and for any function $v(\cdot)$ of vector $[x, y, z, a, b]$,

$$g \circ v(x, y, z, a, b) \triangleq \frac{1}{\left(\frac{(2-y)z}{(1-y)^2} - x\right)v(x, y, z, a, b) + 1}. \quad (10)$$

Under assumption A2, we have the following lower and upper bounds on the throughput U when $M = 1$.

- *Case 1:* $p_{11} \geq p_{01}$

$$\frac{f(c_1)(1 - \epsilon)}{1 - (p_{11} - f(c_1))(1 - \epsilon)} \leq U \leq \frac{\omega_o(1 - \epsilon)}{1 - (p_{11} - \omega_o)(1 - \epsilon)}, \quad (11)$$

where ω_o is given by (2) and

$$\begin{aligned} c_1 &= (\omega_o - c_2)(p_{11} - p_{01})^{N-1}, \\ c_2 &= \frac{p_{01}(1 - p_{01} + \epsilon p_{11})}{1 - p_{01} + \epsilon p_{01}}. \end{aligned}$$

- *Case 2:* $p_{11} < p_{01}$

$$g \circ h(x_1, y_1, z_1, a_1, 2N - 4) \leq U \leq g \circ h\left(\frac{1}{x_1}, 1 - z_1, 1 - y_1, a_1, 3\right), \quad (12)$$

where

$$\begin{aligned} x_1 &= \frac{p_{01}}{p_{11}(p_{11} - p_{01}) + p_{01}}, \\ y_1 &= 1 - (1 - \epsilon)(p_{11}(p_{11} - p_{01}) + p_{01}), \\ z_1 &= (1 - \epsilon)p_{01}, \\ a_1 &= (1 - \epsilon)(\omega_o - p_{11})(p_{11} - p_{01}). \end{aligned}$$

Proof:

- *Case 1:* $p_{11} \geq p_{01}$

Let ω_k denote the belief value of the chosen channel in the first slot of the k -th TP. The length $L_k(\omega_k)$ of this TP has the following distribution.

$$\Pr[L_k(\omega_k) = l] = \begin{cases} 1 - \omega_k(1 - \epsilon), & l = 1 \\ \omega_k(1 - \epsilon)^{k-1} p_{11}^{l-2} (1 - p_{11}(1 - \epsilon)), & l > 1 \end{cases}. \quad (13)$$

It is easy to see that if $\omega' \geq \omega$, then $L_k(\omega')$ stochastically dominates $L_k(\omega)$.

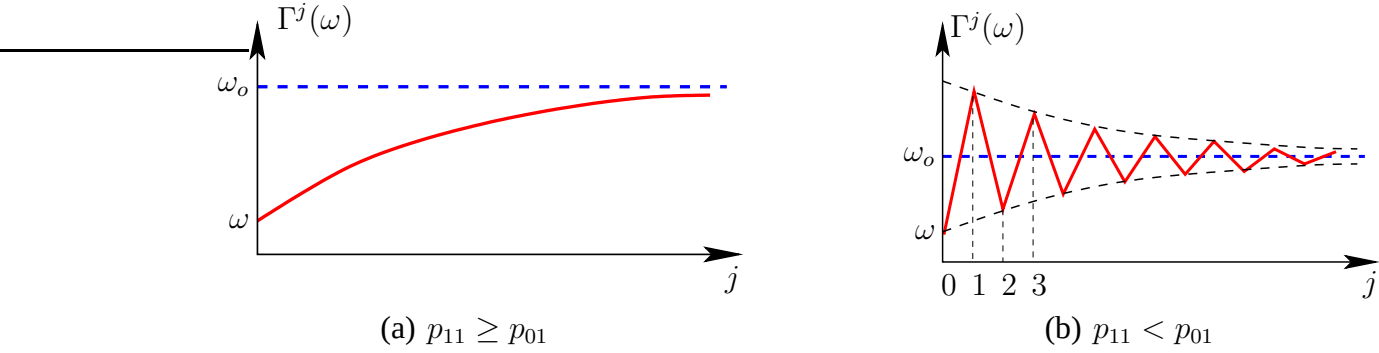


Fig. 3. The j -step belief update when unobserved.

Note that the j -step belief update $\Gamma^j(\omega)$ when unobserved is given by (See Fig. 3)

$$\Gamma^j(\omega) = \omega_o - (\omega_o - \omega)(p_{11} - p_{01})^j.$$

Based on the structure of the myopic policy, we have $\omega_k = \Gamma^{(J_k+1)}(\frac{\epsilon x}{\epsilon x + 1 - x})$, where $J_k = \sum_{i=1}^{N-1} L_{k-i}$ denotes the number of consecutive slots in which the chosen channel has been unobserved since the last visit, and x denotes the belief value of the chosen channel at the last time the user left it. From assumption A2, $\Gamma(\frac{\epsilon x}{\epsilon x + 1 - x}) \leq \Gamma(p_{01}) \leq \omega_o$, where ω_o is the stationary distribution of the Gilbert-Elliot channel given in (2). Based on the monotonic convergence property of the j -step belief update (see Fig. 3 (a)), we have $\omega_k \leq \omega_o$. $L_k(\omega_o)$ thus stochastically dominates $L_k(\omega_k)$, and the expectation of the former, $\overline{L_k(\omega_o)} = 1 + \frac{\omega_o(1-\epsilon)}{1-p_{11}(1-\epsilon)}$, leads to the upper bound of U given in (11).

Next, we prove the lower bound of U by constructing a hypothetical system where the initial belief value of the chosen channel in a TP is a lower bound of that in the real system. The average TP length in this hypothetical system is thus smaller than that in the real system, leading to a lower bound on U based on (6). Specifically, since $\omega_k = \Gamma^{(J_k+1)}(\frac{\epsilon x}{\epsilon x + 1 - x})$ and $J_k = \sum_{i=1}^{N-1} L_{k-i} \geq N + L_{k-1} - 2$, we have $\omega_k \geq \Gamma^{N+L_{k-1}-1}(\frac{\epsilon x}{\epsilon x + 1 - x}) \geq \Gamma^{N+L_{k-1}-1}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}})$

based on the monotonic increasing property of the j -step belief update (see Fig. 3 (a)). We thus construct a hypothetical system given by a first-order Markov chain $\{L'_k\}_{k=1}^\infty$ with the following transition matrix $\mathbf{R} = \{r_{i,j}\}$.

$$r_{i,j} = \begin{cases} 1 - \Gamma^{N+i-1}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}), & i \geq 1, j = 1 \\ \Gamma^{N+i-1}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}})(1 - \epsilon)^{j-1}(p_{11})^{j-2}(1 - p_{11}(1 - \epsilon)), & i \geq 1, j \geq 2 \end{cases}. \quad (14)$$

Lemma 2: The stationary distribution of the first order Markov chain $\{L'_k\}_{k=1}^\infty$ is stochastically dominated by the stationary distribution of $\{L_k\}_{k=1}^\infty$.

Proof:

Let ω'_k denote the expected probability that the chosen channel is in state 1 in the first slot of the k -th transmission period of $\{L'_k\}_{k=1}^\infty$. Assume in the k -th transmission period, the distributions of L'_k and L_k both equal to the same distribution $\vec{\lambda}$, which may or may not be the stationary distribution of $\{L_k\}_{k=1}^\infty$. Next we show $\omega_{k+n} \geq \omega'_{k+n}$ for any $n \geq 1$ by induction.

When $n = 1$, we have

$$\begin{aligned} \omega_{k+1} &= \sum_{l=1}^\infty \mathbb{E}_{L_{k-N+2}, \dots, L_{k-1}} [\Gamma^{1+\sum_{i=k-N+2}^k L_i}(\frac{\epsilon x}{\epsilon x + 1 - x}) | L_k = l] Pr(L_k = l) \\ &\geq \sum_{l=1}^\infty \mathbb{E}_{L_{k-N+2}, \dots, L_{k-1}} [\Gamma^{N-1+L_k}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) | L_k = l] Pr(L_k = l) \\ &= \sum_{l=1}^\infty \Gamma^{N-1+l}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) \lambda_l \\ &= \omega'_{k+1}. \end{aligned} \quad (15)$$

Assume $\omega_{k+n} \geq \omega'_{k+n}$, then

$$\begin{aligned} \omega_{k+n+1} &= \sum_{l=1}^\infty \mathbb{E}_{L_{k+n-N+2}, \dots, L_{k+n-1}} [\Gamma^{1+\sum_{i=k+n-N+2}^{k+n} L_i}(\frac{\epsilon x}{\epsilon x + 1 - x}) | L_{k+n} = l] Pr(L_{k+n} = l) \\ &\geq \sum_{l=1}^\infty \mathbb{E}_{L_{k+n-N+2}, \dots, L_{k+n-1}} [\Gamma^{N-1+L_{k+n}}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) | L_{k+n} = l] Pr(L_{k+n} = l) \\ &= \sum_{l=1}^\infty \Gamma^{N-1+l}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) Pr(L_{k+n} = l) \end{aligned} \quad (16)$$

Since $\omega_{k+n} \geq \omega'_{k+n}$, by (13), we have

$$\begin{aligned} \Pr(L_{k+n} = l) &\leq \Pr(L'_{k+n} = l), \quad \text{if } l = 1; \\ \Pr(L_{k+n} = l) &\geq \Pr(L'_{k+n} = l), \quad \text{if } l > 1. \end{aligned} \quad (17)$$

Since the smallest number in the series $\Gamma^{N-1+l}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}})$ is the first one, by (17) and the fact that $\sum_{l=1}^{\infty} \Pr(L_{k+n} = l) = \sum_{l=1}^{\infty} \Pr(L'_{k+n} = l) = 1$, we have

$$\sum_{l=1}^{\infty} \Gamma^{N-1+l}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) \Pr(L_{k+n} = l) \geq \sum_{l=1}^{\infty} \Gamma^{N-1+l}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}) \Pr(L'_{k+n} = l) = \omega'_{k+n+1} \quad (18)$$

Combine (16) and (18), we have $\omega_{k+n+1} \geq \omega'_{k+n+1}$.

By the above induction, we have $\omega_{k+n} \geq \omega'_{k+n}$ for any $n \geq 1$. So the stationary distribution of the first order Markov chain $\{L'_k\}_{k=1}^{\infty}$ is dominated by the stationary distribution of $\{L_k\}_{k=1}^{\infty}$. ■

Let $\overline{L'}$ denote the average length of a transmission period of L'_k . Based on (6) and Lemma 2, $\overline{L'}$ leads to a lower bound on U . Last, we obtain closed-form $\overline{L'}$ by solving the stationary distribution of the first-order Markov chain $\{L'_k\}_{k=1}^{\infty}$.

Recall that $\mathbf{R} = \{r_{i,j}\}$ is the transition matrix of $\{L_k\}_{k=1}^{\infty}$, where $r_{i,j}$ is given in (14). Let $\mathbf{R}(:, k)$ denote the k -th column of $\mathbf{R}$. We have

$$\mathbf{1} - \mathbf{R}(:, 1) = \frac{\mathbf{R}(:, 2)}{1 - p_{11}(1 - \epsilon)}, \quad \mathbf{R}(:, k) = \mathbf{R}(:, 2)(p_{11}(1 - \epsilon))^{k-2}, \quad (k \geq 2) \quad (19)$$

where $\mathbf{1}$ is the unit column vector $[1, 1, \dots]^t$. By the definition of stationary distribution, we have, for $k = 1, 2, \dots$,

$$[\lambda_1, \lambda_2, \dots] \mathbf{R}(:, k) = \lambda_k, \quad (20)$$

which, combined with (19), leads to

$$\lambda_1 = 1 - \frac{\lambda_2}{(1 - p_{11}(1 - \epsilon))}, \quad \lambda_k = \lambda_2(p_{11}(1 - \epsilon))^{k-2}, \quad (k \geq 2) \quad (21)$$

Substituting (21) into (20) for $k = 2$ and solving for λ_2 , we have $\lambda_2 = f(c_1)(1 - \epsilon)(1 - p_{11}(1 - \epsilon))$, where $f(c_1)$ is given in (11). From (21), we then have the stationary distribution as

$$\lambda_k = \begin{cases} 1 - f(c_1)(1 - \epsilon), & k = 1 \\ f(c_1)(1 - \epsilon)(p_{11}(1 - \epsilon))^{k-2}(1 - p_{11}(1 - \epsilon)), & k > 1 \end{cases}, \quad (22)$$

which leads to $\bar{L} = \sum_{k=1}^{\infty} k \lambda_k = 1 + \frac{f(c_1)(1 - \epsilon)}{1 - p_{11}(1 - \epsilon)}$.

- *Case 2: $p_{11} < p_{01}$*

Let ω_k denote the belief value of the chosen channel in the first slot of the k -th TP. Define the operator $c(\cdot)$ as $c(x) = \frac{\epsilon x}{\epsilon x + 1 - x}$. We have

$$\Pr[L_k(\omega_k) = l] = \begin{cases} \omega_k(1 - \epsilon), & l = 1 \\ (1 - \omega_k(1 - \epsilon)) \prod_{i=1}^{l-2} (1 - (\Gamma \circ c)^i(\omega_k)(1 - \epsilon)) (\Gamma \circ c)^{l-1}(\omega_k)(1 - \epsilon), & l > 1 \end{cases} \quad (23)$$

Consider first the upper bound. We construct the following hypothetical system where the stationary distribution of a TP is stochastically dominated by the one in the real system. The average TP length in this hypothetical system is thus smaller than in the real system, leading to an upper bound on U based on (6). Specifically, the distribution of a TP in the hypothetical system has the following form.

$$\Pr[L'_k(\omega_k) = l] = \begin{cases} \frac{\Gamma(p_{11})}{p_{01}} \omega_k(1 - \epsilon) + 1 - \frac{\Gamma(p_{11})}{p_{01}}, & l = 1 \\ (1 - \omega_k(1 - \epsilon))(1 - p_{01}(1 - \epsilon))^{k-2} \Gamma(p_{11})(1 - \epsilon), & l > 1 \end{cases} \quad (24)$$

We first show that $L'_k(\omega_k)$ is stochastically dominated by $L_k(\omega_k)$. Note that $\Pr[L'_k(\omega_k) = l] \geq 0$ for all $l \in \mathcal{Z}^+$ and $\sum_{l=1}^{\infty} \Pr[L'_k(\omega_k) = l] = 1$. The distribution of $L'_k(\omega_k)$ given in (24) is thus well-defined. Since $\Gamma(p_{11}) \leq \Gamma \circ c(\omega) \leq p_{01}$ for any $p_{11} \leq \omega \leq p_{01}$, we have $\Pr[L'_k(\omega_k) = l] \leq \Pr[L_k(\omega_k) = l]$ for all $l \geq 2$. $L'_k(\omega_k)$ is thus stochastically dominated by $L_k(\omega_k)$.

It is easy to see that $L'_k(\omega')$ is stochastically dominated by $L'_k(\omega)$ if $\omega' \geq \omega$. $L'_k(\omega')$ is thus stochastically dominated by $L_k(\omega)$ if $\omega' \geq \omega$. Based on the structure of the myopic policy, it is clear that when L_{k-1} is odd, in the k -th TP, the user will switch to the channel visited in the $(k-2)$ -th TP. As a consequence, the initial belief ω_k of the k -th TP is given by $\omega_k = \Gamma^{(L_{k-1}+1)}(1)$. When L_{k-1} is even, we can show that $\omega_k \leq \Gamma^{(L_{k-1}+4)}(1)$. This is because that for L_{k-1} even, the user cannot switch to a channel visited $L_{k-1} + 2$ slots ago, and $\Gamma^j(1)$ decreases with j for even j 's and $\Gamma^j(1) \geq \Gamma^i(1)$ for any even j and odd i (see Fig. 3 (b)). We thus construct a hypothetical system given by the first-order Markov chain $\{L'_k\}_{k=1}^{\infty}$ with the following transition probabilities.

$$r_{i,j} = \begin{cases} \frac{\Gamma(p_{11})}{p_{01}} \Gamma^{i+1}(1)(1 - \epsilon) + 1 - \frac{\Gamma(p_{11})}{p_{01}}, & \text{if } i \text{ is odd, } j = 1 \\ (1 - \Gamma^{i+1}(1)(1 - \epsilon))(1 - p_{01}(1 - \epsilon))^{j-2} \Gamma(p_{11})(1 - \epsilon), & \text{if } i \text{ is odd, } j \geq 2 \\ \frac{\Gamma(p_{11})}{p_{01}} \Gamma^{i+4}(1)(1 - \epsilon) + 1 - \frac{\Gamma(p_{11})}{p_{01}}, & \text{if } i \text{ is even, } j = 1 \\ (1 - \Gamma^{i+4}(1)(1 - \epsilon))(1 - p_{01}(1 - \epsilon))^{j-2} \Gamma(p_{11})(1 - \epsilon), & \text{if } i \text{ is even, } j \geq 2 \end{cases}.$$

Similarly to Lemma 2, it can be shown that the stationary distribution of $\{L'_k\}_{k=1}^\infty$ is stochastically dominated by that of $\{L_k\}_{k=1}^\infty$. Furthermore the stationary distribution of $\{L'_k\}_{k=1}^\infty$ can be obtained in closed form by using an approach similar to that in Case 1, leading to the upper bound on U given in (12).

We now prove the lower bound. Consider the hypothetical system with the distribution of a TP as given below.

$$\Pr[L'_k(\omega_k) = l] = \begin{cases} \frac{p_{01}}{\Gamma(p_{11})}\omega_k(1-\epsilon) + 1 - \frac{p_{01}}{\Gamma(p_{11})}, & l = 1 \\ (1 - \omega_k(1-\epsilon))(1 - \Gamma(p_{11})(1-\epsilon))^{k-2}p_{01}(1-\epsilon), & l > 1 \end{cases}. \quad (25)$$

Similarly, $L'_k(\omega_k)$ is well-defined and stochastically dominates $L_k(\omega_k)$. It is easy to see that $L'_k(\omega')$ stochastically dominates $L'_k(\omega)$ if $\omega' \leq \omega$. $L'_k(\omega')$ thus stochastically dominates $L_k(\omega)$ if $\omega' \leq \omega$.

Based on the structure of the myopic policy, $\omega_k = p_{11}^{(L_{k-1}+1)}$ when L_{k-1} is odd. When L_{k-1} is even, to find a lower bound on ω_k , we need to find the smallest odd j such that the last visit to the channel chosen in the k -th TP is j slots ago. From the structure of the myopic policy, the smallest feasible odd j is $L_{k-1} + 2N - 3$, which corresponds to the scenario where all N channels are visited in turn from the $(k - N + 1)$ -th TP to the k -th TP with $L_{k-N+1} = L_{k-N+2} = \dots = L_{k-2} = 2$. We thus have $\omega_k \geq p_{11}^{(L_{k-1}+2N-3)}$. We then construct a hypothetical system given by the first-order Markov chain $\{L'_k\}_{k=1}^\infty$ with the following transition probabilities.

$$r_{i,j} = \begin{cases} \frac{p_{01}}{\Gamma(p_{11})}\Gamma^{i+1}(1)(1-\epsilon) + 1 - \frac{p_{01}}{\Gamma(p_{11})}, & \text{if } i \text{ is odd, } j = 1 \\ (1 - \Gamma^{i+1}(1)(1-\epsilon))(1 - p_{01}(1-\epsilon))^{j-2}\Gamma(p_{11})(1-\epsilon), & \text{if } i \text{ is odd, } j \geq 2 \\ \frac{\Gamma(p_{11})}{p_{01}}\Gamma^{i+2N-3}(1)(1-\epsilon) + 1 - \frac{\Gamma(p_{11})}{p_{01}}, & \text{if } i \text{ is even, } j = 1 \\ (1 - \Gamma^{i+2N-3}(1)(1-\epsilon))(1 - p_{01}(1-\epsilon))^{j-2}\Gamma(p_{11})(1-\epsilon), & \text{if } i \text{ is even, } j \geq 2 \end{cases}.$$

The stationary distribution of this hypothetical system leads to the lower bound on U given in (12). ■

For multi-channel sensing ($M > 1$), it is difficult to construct first-order Markov process to stochastically dominate or be dominated by $\{L_k\}_{k=1}^\infty$. Instead, we establish a uniform statistical bound on the distributions of all TPs based on the structure of the myopic policy. The bounds on the throughput when applied $M = 1$ are thus looser than those under single-channel sensing scenarios as given in Theorem 2.

Theorem 3: Recall the definition of $g \circ v(\cdot)$ given in (10). Under assumption A2, we have the following lower and upper bounds on throughput U when $M > 1$.

- *Case 1:* $p_{11} \geq p_{01}$

$$\frac{Mc_3(1-\epsilon)}{1-(p_{11}-c_3)(1-\epsilon)} \leq U \leq \frac{M\omega_o(1-\epsilon)}{1-(p_{11}-\omega_o)(1-\epsilon)}. \quad (26)$$

where $c_3 = \omega_o - (\omega_o - \frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}})(p_{11} - p_{01})^{\lfloor \frac{N}{M} \rfloor}$.

- *Case 2:* $p_{11} < p_{01}$

$$Mg \circ v_1(x_1, y_1, z_1, a_1, b_1) \leq U \leq Mg \circ v_2(\frac{1}{x_1}, 1 - z_1, 1 - y_1, a_1, b_1), \quad (27)$$

where

$$\begin{aligned} v_1(\cdot) &= 1 - (\omega_o - (\omega_o - p_{11})(p_{11} - p_{01})^{2\lfloor \frac{N}{M} \rfloor - 2})(1 - \epsilon), \\ v_2(\cdot) &= 1 - (p_{11}(p_{11} - p_{01}) + p_{01})(1 - \epsilon), \\ x_1 &= \frac{p_{01}}{p_{11}(p_{11} - p_{01}) + p_{01}}, \\ y_1 &= 1 - (p_{11}(p_{11} - p_{01}) + p_{01})(1 - \epsilon), \\ z_1 &= (1 - \epsilon)p_{01}. \end{aligned}$$

Note that a_1 and b_1 can be arbitrary since they are arguments of the constant functions v_1 and v_2 .

Proof:

- *Case 1:* $p_{11} \geq p_{01}$

Consider first the upper bound. Similarly to single-channel sensing, the belief value ω_k of the chosen channel in the first slot of the k -th TP is upper bounded by ω_o . $L_k(\omega_o)$ thus stochastically dominates $L_k(\omega_k)$, and the expectation of the former leads to the upper bound on U given in (26).

We now consider the lower bound. Recall that $\omega_k = \Gamma^{(J_k+1)}(\frac{\epsilon x}{\epsilon x + 1 - x})$, where J_k denotes the number of consecutive slots in which the chosen channel has been unobserved since the last visit, and x denotes the belief value of the chosen channel at the last time the user left it. Based on the structure of the myopic policy, the channel has the last priority when the user leaves it. It will take at least $\lfloor \frac{N-M}{M} \rfloor$ slots before the user returns to the same channel, *i.e.*, $J_k \geq \lfloor \frac{N}{M} \rfloor - 1$. Based on the monotonic increasing property of the j -step transition probability $\mathcal{T}^j(\omega)$ (see Fig. 3 (a)), we have $\omega_k = \Gamma^{J_k+1}(\frac{\epsilon x}{\epsilon x + 1 - x}) \geq \Gamma^{\lfloor \frac{N}{M} \rfloor}(\frac{\epsilon x}{\epsilon x + 1 - x}) \geq \Gamma^{\lfloor \frac{N}{M} \rfloor}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}})$. Thus $L_k(\Gamma^{\lfloor \frac{N}{M} \rfloor}(\frac{\epsilon p_{01}}{\epsilon p_{01} + 1 - p_{01}}))$

is stochastically dominated by $L_k(\omega_k)$, and the expectation of the former leads to the lower bound on U given in (26).

- *Case 2:* $p_{11} < p_{01}$

Consider first the upper bound. Let ω_k denote the belief value of the chosen channel in the first slot of the k -th TP. Based on the structure of the myopic policy, we have $\omega_k = \Gamma^{J_k+1}(1)$, where J_k denotes the number of consecutive slots in which the chosen channel has been unobserved since the last visit. From Fig. 3 (b), we have $\omega_k = \Gamma^{J_k+1}(1) \leq \Gamma^2(1)$. Combined with the hypothetical system given in (24), $L'_k(\Gamma^2(1))$ is stochastically dominated by $L_k(\omega_k)$, and the expectation of the former leads to the upper bound on U given in (27).

We now consider the lower bound. Recall that $\omega_k = \Gamma^{J_k+1}(1)$. If J_k is odd, then $\Gamma^{J_k+1}(1) \geq \Gamma^{2\lfloor \frac{N}{M} \rfloor - 1}(1)$ since $2\lfloor \frac{N}{M} \rfloor - 1$ is an odd number (see Fig. 3 (b)). If J_k is even, i.e., the user has stayed even slots before it returns this channel, then J_k is at least $2\lfloor \frac{N-M}{M} \rfloor$. we have $\omega_k = \Gamma^{J_k+1}(1) \geq \Gamma^{2\lfloor \frac{N}{M} \rfloor - 1}(1)$. Combined with the hypothetical system given in (25), $L'_k(\Gamma^{2\lfloor \frac{N}{M} \rfloor - 1}(1))$ stochastically dominates $L_k(\omega_k)$, and the expectation of the former leads to the lower bound on U given in (27). ■

Corollary 2: For $p_{11} > p_{01}$, the lower bound on throughput U increasingly converges to the constant upper bound at geometrical rate $(p_{11} - p_{01})^{\frac{1}{M}}$ as N increases; for $p_{11} < p_{01}$, the lower bound on U increasingly converges to a constant at geometrical rate $(p_{01} - p_{11})^{\frac{2}{M}}$.

Proof: From the closed-form expressions of the lower bounds on U given in Theorem 2 and Theorem 3, it is easy to see that the lower bound is monotonically increasing with N . Let $x = |p_{11} - p_{01}|$. For $p_{11} > p_{01}$, after some simplifications, the lower bound has the form $a + b/(x^{\lfloor \frac{N}{M} \rfloor} + c)$, where a, b, c ($c \neq 0$) are constants. The upper bound is $a + b/c$. We have $\frac{|a+b/(x^{\lfloor \frac{N}{M} \rfloor} + c) - a - b/c|}{x^{\frac{N}{M}}} \rightarrow O(b/c^2)$ as $N \rightarrow \infty$. Thus the lower bound converges to the upper bound with geometric rate $x^{\frac{1}{M}}$.

For $p_{11} < p_{01}$, the lower bound has the form $d + e/(x^{2\lfloor \frac{N}{M} \rfloor - 1} + f)$, where d, e, f ($f \neq 0$) are constants. It converges to $d + e/f$ as $N \rightarrow \infty$. We have $\frac{|d+e/(x^{2\lfloor \frac{N}{M} \rfloor - 1} + f) - d - e/f|}{x^{\frac{2N}{M}}} \rightarrow O(e/(xf^2))$ as $N \rightarrow \infty$. Thus the lower bound converges with geometric rate $x^{\frac{2}{M}}$. ■

The convergence of the lower bound to the upper bound when $p_{11} \geq p_{01}$ can be explained as follows. The upper bound given in Theorem 3 corresponds to the case where the belief in the first slot of a TP is equal to the stationary distribution ω_o . If a user can always switch to channels

with probability ω_o being in good state when channel switches are needed, the throughput will achieve the upper bound given in (26). Specifically, we have the following theorem, which gives the closed-form performance under the myopic policy over a finite horizon.

Theorem 4: For $p_{11} \geq p_{01}$ and $N \geq MT$, under assumption A2, the expected total reward over T slots when the initial belief starts from the stationary distribution is given below.

$$\hat{V}_{1:T}(\Omega(1)) = M\omega_o(1 - \epsilon) \left(\frac{(T - 1)}{1 - (p_{11} - \omega_o)(1 - \epsilon)} + c_4 \right), \quad (28)$$

where

$$c_4 = 1 - \frac{((\omega_o - p_{11})(1 - \epsilon))^2 (1 - ((p_{11} - \omega_o)(1 - \epsilon))^{T-1})}{(1 - (p_{11} - \omega_o)(1 - \epsilon))^2}.$$

Proof: From the structure of the myopic policy, if the user observes state 1 from a channel, it will stay on that channel. Otherwise, it will switch to a new channel (with belief ω_o). Clearly, V does not depend on N since at most MT channels need to be considered during T slots.

In the first slot, the user randomly chooses M channels and gets $M\omega_o(1 - \epsilon)$ units of reward. Then the user will either stay or switch on a channel. This process is a Markov chain with states “stay” and “switch” as shown below.

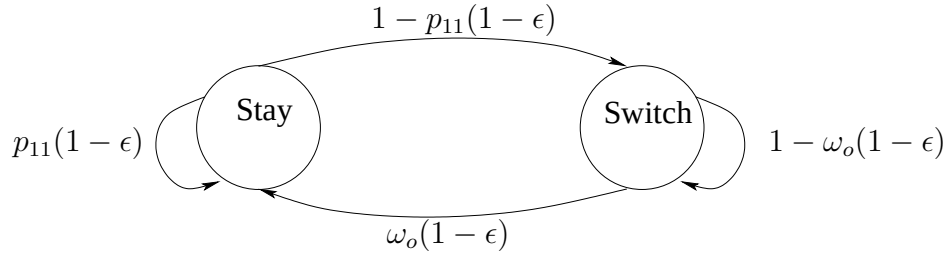


Fig. 4. The Markov chain with states “stay” and “switch”.

If the user observes 1 on a channel after the first slot, it will stay and get $p_{11}(1 - \epsilon)$ units of reward on this channel. Otherwise it will switch to a new channel and get $\omega_o(1 - \epsilon)$ units of reward. So V is determined by the distribution of the states of the above two-state Markov chain.

$$\begin{aligned}
V(\omega_o, T) &= M(\Sigma_{M=1}^{T-1} [\omega_o(1-\epsilon) \quad 1-\omega_o(1-\epsilon)] \begin{bmatrix} p_{11}(1-\epsilon) & 1-p_{11}(1-\epsilon) \\ \omega_o(1-\epsilon) & 1-\omega_o(1-\epsilon) \end{bmatrix}^{M-1} \begin{bmatrix} p_{11}(1-\epsilon) \\ \omega_o(1-\epsilon) \end{bmatrix} + \omega_o(1-\epsilon)) \\
&= M(\Sigma_{M=1}^{T-1} [\omega_o(1-\epsilon) \quad 1-\omega_o(1-\epsilon)] \{ \frac{1}{1-(p_{11}-\omega_o)(1-\epsilon)} \begin{bmatrix} \omega_o(1-\epsilon) & 1-p_{11}(1-\epsilon) \\ \omega_o(1-\epsilon) & 1-p_{11}(1-\epsilon) \end{bmatrix} \\
&\quad + \frac{((p_{11}-\omega_o)(1-\epsilon))^{M-1}}{1-(p_{11}-\omega_o)(1-\epsilon)} \begin{bmatrix} 1-p_{11}(1-\epsilon) & p_{11}(1-\epsilon)-1 \\ -\omega_o(1-\epsilon) & \omega_o(1-\epsilon) \end{bmatrix} \} \begin{bmatrix} p_{11}(1-\epsilon) \\ \omega_o(1-\epsilon) \end{bmatrix} + \omega_o(1-\epsilon)) \\
&= M(\frac{\omega_o(1-\epsilon)(T-1)}{1-(p_{11}-\omega_o)(1-\epsilon)} - \frac{\omega_o(1-\epsilon)^3(\omega_o-p_{11})^2(1-((p_{11}-\omega_o)(1-\epsilon))^{T-1})}{(1-(p_{11}-\omega_o)(1-\epsilon))^2} + \omega_o(1-\epsilon))
\end{aligned} \tag{29}$$

■

From (29), we immediately see that the throughput U is given as follows.

$$U = \lim_{T \rightarrow \infty} \frac{\hat{V}_{1:T}(\Omega(1))}{T} = \frac{M\omega_o(1-\epsilon)}{1-(p_{11}-\omega_o)(1-\epsilon)}, \tag{30}$$

which agrees with the upper bound given in Theorem 3.

The monotonicity of the difference between the upper and lower bounds with respect to N illustrates that the performance of the multi-channel opportunistic system improves with the number N of channels, as suggested by intuition. For $p_{11} \geq p_{01}$, the upper bound gives the limiting performance of the opportunistic system when $N \rightarrow \infty$. By Corollary 2, the throughput of a multi-channel opportunistic system with single-channel sensing quickly saturates as the number of channels increases; it is thus crucial to enhance radio sensing capability in order to fully exploit the communication opportunities offered by a large number of channels.

IV. APPROXIMATION FACTOR OF THE MYOPIC POLICY

Although the optimality of the myopic policy is proved for $N = 2$ and conjectured for general scenarios based on numerical results, establishing the optimality or simple sufficient conditions for optimality appears to be challenging. Under the discounted reward criterion, we have shown that so long as the discount factor is less than $1/(M+1)$, the myopic policy is optimal for all N . In this section, we take a further step toward the optimality of the myopic policy. By considering a genie aided system, we establish a bound on the performance loss of the myopic policy and its approximation factor regarding to the optimal policy.

A. A Genie-aided System

In the Genie-aided system, we assume the user can sense, access, and obtain observations (ACK/NAK) from all N channels. However, the user can only get reward from M channels determined at the beginning of each slot. Clearly, the myopic policy (*i.e.*, choose M channels with largest probabilities of being in state 1 to accrue rewards) is optimal since current choice will not affect the belief transitions. Similar to Corollary 1, the reward process of the genie-aided system is ergodic under assumption A2. Furthermore, we obtain an upper bound on the optimal performance of the genie-aided system.

Theorem 5: Define $x \triangleq \frac{\epsilon p_{11}^2 + p_{01} - p_{01} p_{11}}{\epsilon p_{11} + 1 - p_{11}}$. Under assumption A2, the maximum steady-state throughput $\bar{U}$ in the genie-aided system is upper bounded as given below.

- *Case 1:* $p_{11} \geq p_{01}$

$$\bar{U} \leq (M p_{11} - \sum_{k=0}^M \binom{N}{k} d_k) (1 - \epsilon), \quad (31)$$

where

$$d_k = (M - k)(p_{11} - x)(\omega_o(1 - \epsilon))^k (1 - \omega_o(1 - \epsilon))^{N-k}.$$

- *Case 2:* $p_{11} < p_{01}$

$$\bar{U} \leq (M x - \sum_{k=0}^M \binom{N}{k} e_k) (1 - \epsilon), \quad (32)$$

where

$$e_k = (M - k)(x - p_{11})(\omega_o(1 - \epsilon))^{N-k} (1 - \omega_o(1 - \epsilon))^k.$$

Proof:

- *Case 1:* $p_{11} \geq p_{01}$

Based on the ergodicity of the reward process in the genie-aided system, the initial belief vector does not affect the optimal performance. Without loss of generality, assume the state of each channel starts from the stationary distribution ω_o . As a consequence, the number k of channels observed as 1 falls into the binomial distribution $B(k, N, \omega_o(1 - \epsilon))$ in every slot. Since the channels observed as 1 will have the largest belief value p_{11} and other channels' belief values will be upper bounded by $\Gamma(\frac{\epsilon p_{11}}{\epsilon p_{11} + 1 - p_{11}})$ in the next slot, the expected reward obtained under the

myopic policy will be upper bound by the right-hand side of (31). We thus proved the upper bound on $\bar{U}$.

- *Case 2:* $p_{11} < p_{01}$

Similarly, we assume the state of each channel starts from the stationary distribution ω_o without loss of generality. The number k of channels observed as 1 falls into the binomial distribution $B(k, N, \omega_o(1 - \epsilon))$ in every slot. Since the channels observed as 1 will have the smallest belief value p_{11} and other channels' belief values will be upper bounded by $\Gamma(\frac{\epsilon p_{11}}{\epsilon p_{11} + 1 - p_{11}})$ in the next slot, the expected reward obtained under the myopic policy will be upper bound by right-hand side of (32). We thus proved the upper bound on $\bar{U}$. ■

B. Approximation Factor

Clearly, the optimal performance of the genie-aided system is an upper bound on the maximum throughput in the original multi-channel opportunistic access system. In other words, $\bar{U}$ provides a performance benchmark of all sensing policies, including the myopic policy. To better bound the performance of the myopic policy, we present another lower bound on the throughput U under the myopic policy.

Theorem 6: Let $\tilde{U}$ be the throughput under random sensing policy that chooses M out of N channels with uniform probability (*i.e.*, choose any set of M channels with probability $1/\binom{N}{M}$), and U^* the maximum throughput under the optimal policy. We have

$$M\omega_o(1 - \epsilon) = \tilde{U} \leq U \leq U^* \leq \bar{U} \leq N\omega_o(1 - \epsilon). \quad (33)$$

Proof: Since channels are stochastically identical, the random sensing policy is equivalent to the static policy that chooses a constant set of M channels in each slot. Clearly, the long-run throughput of the static policy on a chosen channel is given by the stationary distribution ω_o multiplied by the probability $(1 - \epsilon)$ of no false alarm.

To prove $\tilde{U} \leq U$, we note that the expected immediate reward under the random sensing policy in each slot is given by the expected sum of M randomly chosen belief values under any given policy (including the myopic policy). Since the expected immediate reward under the myopic policy in each slot is given by the expected sum of the first M largest belief values. The throughput under the myopic policy is thus lower bounded by that under the random sensing policy.

The proof for $U \leq U^* \leq \bar{U}$ is trivial. To prove $\bar{U} \leq N\omega_o(1 - \epsilon)$, we note that $N\omega_o(1 - \epsilon)$ is the throughput under the policy that senses and accrues rewards from all of the N channels. ■

Combining the maximum of the lower bounds on U given in Theorem 2, Theorem 3 and Theorem 6 and the minimum of the upper bounds on $\bar{U}$ given in Theorem 5 and Theorem 6, we obtain a uniform bound on the performance loss under the myopic policy. We further obtain the approximation factor of the myopic policy as given below.

Corollary 3: Let $\eta \triangleq \frac{U}{U^*}$ ($\eta \in [0, 1]$) be the approximation factor of the myopic policy. Under assumption A2, we have

$$\eta \geq \begin{cases} \frac{M}{N}, & \text{if } p_{11} > p_{01} \\ \max\{\frac{1}{2}, \frac{M}{N}\}, & \text{if } p_{11} < p_{01} \\ 1, & \text{if } p_{11} = p_{01} \end{cases} . \quad (34)$$

Proof: From Theorem 6, we directly see that $\eta \geq \frac{M}{N}$. Consider $p_{11} < p_{01}$. Based on Theorem 5, we have $\bar{U} \leq M\Gamma(\frac{\epsilon p_{11}}{\epsilon p_{11} + 1 - p_{11}})$ (see the proof of Theorem 5). We thus have

$$\eta = \frac{U}{U^*} \geq \frac{\tilde{U}}{\bar{U}} \geq \frac{M\omega_o}{M\Gamma(\frac{\epsilon p_{11}}{\epsilon p_{11} + 1 - p_{11}})} \geq \frac{1}{1 + p_{01} - p_{11}} \geq \frac{1}{2}.$$

For the trivial case $p_{11} = p_{01}$, we note that the lower bound on U given in Theorem 3 agrees with the upper bound on $\bar{U}$ given in Theorem 5. ■

V. NUMERICAL EXAMPLES

In this section, we demonstrate the tightness of the bounds on U given in Sec. III-C2 and Sec. IV-A. In particular, we are interested in the lower and upper bounds on the performance of the myopic policy given in Theorem 2 and Theorem 3, and the upper bound on the optimal performance in the genie-aided system given in Theorem 5. We also generate the performance of the myopic policy and the optimal performance in the genie-aided system by Monte Carlo simulations. Fig. 5 illustrates the bounds on the performance of the myopic policy under single-channel sensing. Fig. 6 illustrates the bounds on the performance of the myopic policy under multi-channel sensing ($M = 2$). We observe that the lower bound on the performance of the myopic policy quickly converges to the upper bound as $N \rightarrow \infty$ when channels are positively correlated. We also observe from Fig. 5–6 that the upper bound on the optimal performance in the genie-aided system is tight.

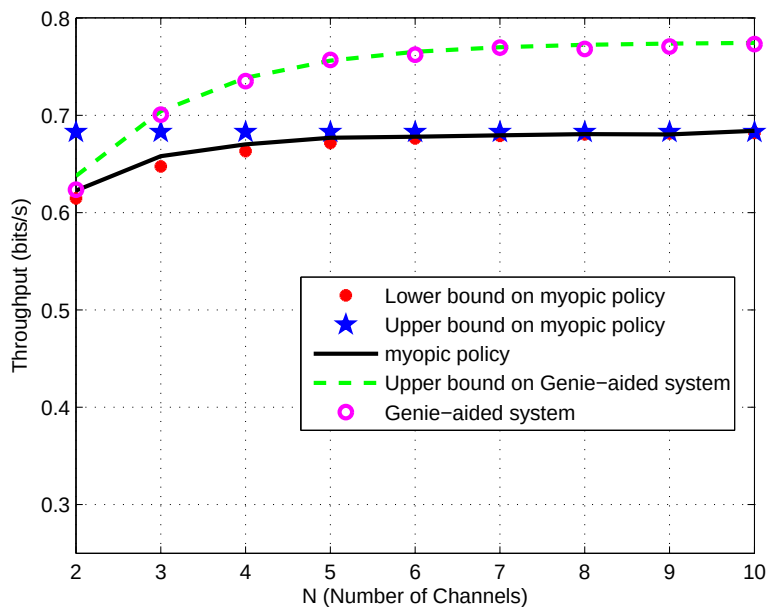


Fig. 5. Performance bounds of the myopic policy ($M = 1$, $p_{11} = 0.8$, $p_{01} = 0.2$, $\epsilon = 0.0312$).

VI. CONCLUSION AND FUTURE WORK

In this paper, we have analyzed the performance of the myopic sensing policy in multi-channel opportunistic access under an independent and stochastically identical Gilbert-Elliot channel model with noisy state observations. Based on the conjectured optimality of the myopic sensing policy, the obtained analytical results allow us to systematically examine the impact of the number of channels and channel dynamics (transition probabilities) on the system performance. An approximation factor of the myopic policy has been established. Future work includes proving the optimality conjecture of the myopic policy, and generalization to independent and stochastically non-identical channel model by investigating Whittle's index policy.

REFERENCES

- [1] E.N. Gilbert, "Capacity of burst-noise channels," *Bell Syst. Tech. J.*, vol. 39, pp. 1253-1265, Sept. 1960. WA), pp. 331-335, June 1995.
- [2] Q. Zhao and B. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing magazine*, vol. 24, pp. 79-89, May 2007.
- [3] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc

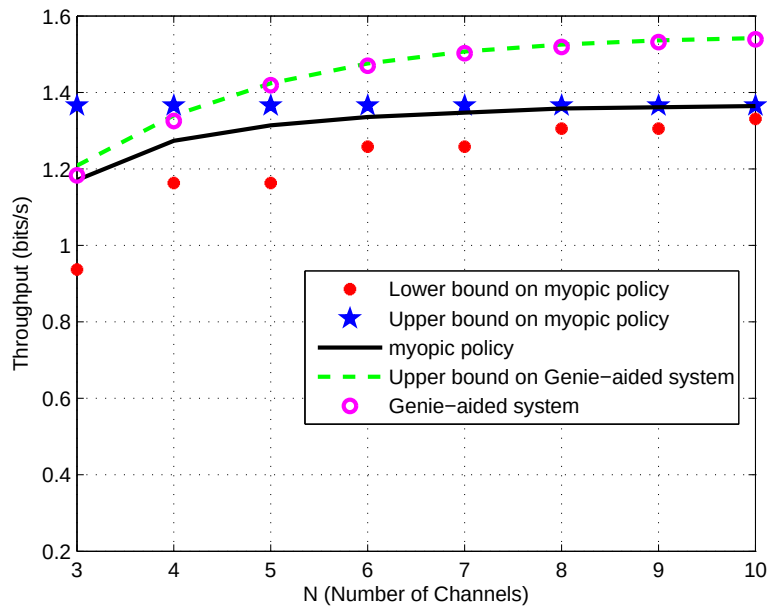


Fig. 6. Performance bounds of the myopic policy ($M = 2$, $p_{11} = 0.8$, $p_{01} = 0.2$, $\epsilon = 0.0312$).

- Networks: A POMDP Framework,” in *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589-600, April, 2007.
- [4] C. H. Papadimitriou and J. N. Tsitsiklis, “The Complexity of Optimal Queueing Network Control,” in *Mathematics of Operations Research*, Vol. 24, No. 2, May 1999, pp. 293-305.
- [5] Q. Zhao, B. Krishnamachari, and K. Liu, “On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance,” in *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431-5440, December, 2008.
- [6] S. H. Ahmad, M. Liu, T. Javadi, Q. Zhao and B. Krishnamachari, “Optimality of Myopic Sensing in Multi-Channel Opportunistic Access,” submitted to *IEEE Transactions on Information Theory*, May, 2008.
- [7] K. Liu and Q. Zhao, “Indexability of Restless Bandit Problems and Optimality of Whittle’s Index for Dynamic Multichannel Access,” submitted to *IEEE Transactions on Information Theory*, November, 2008. Available at <http://arxiv.org/abs/0810.4658> (conference versions appeared in *Proc. of the 5th IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON) Workshops*, June, 2008 and *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, October, 2008).
- [8] Y. Chen, Q. Zhao, and A. Swami, “Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Errors,” in *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053-2071, May, 2008.
- [9] Q. Zhao, B. Krishnamachari, and K. Liu, “Low-Complexity Approaches to Spectrum Opportunity Tracking,” in *Proc. of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, August 2007.
- [10] R. Smallwood and E. Sondik, “The optimal control of partially observable Markov processes over a finite horizon,” *Operations Research*, pp. 1071-1088, 1971.