Assessing the Appropriateness of using Markov Decision Processes for RF Spectrum Management

John Meier, Benjamin Karaus, Sreeharsha Sistla, Terry Tidwell,
Roger D. Chamberlain, and Christopher Gill
Dept. of Computer Science and Engineering
Washington University in St. Louis
{jmeier,bvkaraus,ssistla,terrytidwell,roger,cdgill}@wustl.edu

ABSTRACT
The stochastic nature of wireless communication suggests a Markov Decision Process (MDP) as a formalism for identifying and evaluating spectrum control policies. However, in practice numerous factors influence the success or failure of a transmission, so that the applicability of particular MDP models to real spectrum management problems must itself be examined. This paper presents a series of model validation studies in which correspondence between an MDP model and a discrete-event simulation (DES) model is evaluated. We test several hypotheses that together provide a foundation and an exemplar for the idea of using MDPs to guide management of shared spectrum. We conclude that there is sufficient similarity between the performance predictions made by the MDP model and the DES model that MDPs can be used effectively to determine spectrum control policies.

Categories and Subject Descriptors
C.2.1 [Computer-Communication Networks]: Network Architecture and Design—*Wireless c ommunication*; G.3 [Probability and Statistics]: Markov processes; I.6.4 [Simulation and Modeling]: Model Validation and Analysis

General Terms
Design, Performance, Verification.

Keywords
Markov decision process (MDP); discrete-event simulation (DES); spectrum allocation

1. INTRODUCTION
The advent of software defined radio [1] and techniques for cognitive radio [9] have together opened the door to dramatic improvements in the utilization of scarce RF spectrum [19][23]. Rather than ranges of spectrum being rigidly allocated to specific functions or users, it is now technically feasible for different users (e.g., transmitters) to share available spectrum efficiently, dynamically choosing not only the range of frequencies utilized, but also the modulation method, transmitter power, etc. However, while the technical capability exists to share spectrum effectively, approaches to managing and evaluating dynamic spectrum sharing are still in their infancy.

Anchora et al. [3] have extended the popular ns-3 simulator to enable the assessment of shared spectrum on the part of network operators using game theoretic techniques. At the same time, Feeney [8] cautions that there are concerns about the reliability of simulation results on the part of the wireless community. Proactively assessing the quality of performance predictions is important if the conclusions made on the basis of these predictions are to be trusted. This paper makes an initial effort in that regard for Markov decision processes applied to managing spectrum.

Markov decision processes (MDPs) [16] have been used extensively to optimize control of systems [1][18], particularly those that include stochastic elements [4]. Here, we investigate the viability of using an MDP to optimize decisions in the context of managing RF spectrum. These decisions might be admission decisions, channel allocation decisions, modulation decisions, transmitter power level decisions, or any number of other choices that are germane to managing the shared use of the spectrum.

Critical to the viability of MDP approaches is the question of whether or not the assumptions inherent in the MDP model overly simplify the underlying reality and, as a result, even an optimal choice in the space of the MDP model might or might not be even a good choice in the real world. This paper seeks to assess that question via a series of validation experiments, comparing performance predictions of a basic MDP model to those of a pre-existing discrete-event simulation (DES) model [15]. Note that the MDP will be used to choose a value-optimal policy for spectrum use. However, the DES does not make policy decisions; it only makes performance predictions for a given policy. Our purpose here is to assess the degree to which the MDP model's performance predictions agree with (or differ from) the DES model's performance predictions when they both are following the *same* policy, whether or not that policy is the value-optimal policy chosen by the MDP.

We start by presenting an MDP that models a straightforward admission control problem, one for which we presume to already know the correct allocation decisions. We next demonstrate that this MDP does, in fact, guide us to a correct policy. This is followed by testing a set of hypotheses that examine several

MSWiM'13, November 3–8, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2507924.2507942

aspects of the assumptions made in the MDP modeling process. For each of these hypotheses, a specific experiment is designed and executed, and the validity of the hypothesis is assessed.

The three hypotheses are: (1) performance can be reasonably characterized by the mean of message durations and is relatively insensitive to their distribution; (2) even though imperfect channel allocations will occur in any real system, they are infrequent enough that ignoring them does not have a significant impact on an MDP model's ability to predict throughput accurately; and (3) value-optimal policy decisions made by the MDP are at least locally optimal as determined by the DES model.

The first two hypotheses are assessed by experiments utilizing the discrete-event simulator. The third hypothesis is assessed by comparing the performance predictions of the MDP model to those of the DES model. In all three cases, we find strong evidence to support the hypotheses, providing an indication that Markov decision processes offer a suitable mechanism for modeling and generating policies to manage the shared use of RF spectrum.

The paper is organized as follows. Section 2 describes related work in model validation and Markovian modeling of RF spectrum. Section 3 introduces the RF spectrum system model that we utilize, the MDP model, and their relationship to one another. Section 4 presents the evaluation approach, including a description of the discrete-event simulator we use for validation purposes. Section 5 presents our experimental results, and Section 6 offers conclusions and directions for future work.

2. RELATED WORK

The primary purpose of this paper is to assess the applicability of Markovian models to managing wireless radio spectrum. Perhaps the most relevant results in model applicability come from the domain of finite-element modeling [20], where explicit error estimation is used to select between p-methods and h-methods for analysis. In that work, a quantitative estimate of model error is used to assess whether or not a model is appropriate for a given task.

More commonly, assessment is an empirical exercise, in which the model in question (or more precisely, a set of predictions made by the model) is compared either with measurements of the physical system being modeled (e.g., see [4]) or with a (presumably) more robust model. For example, a frequent practice is to utilize simulation models to assess analytic models [14], as we do in this work. Such an approach makes the most sense when: (1) the simulation model explicitly incorporates aspects of the physical system being modeled that are either simplified or completely ignored in the analytic model, and (2) when the simulation model (or other reference model) has been independently evaluated. To assess the applicability of Markovian models to the problem of managing RF spectrum, we compare model predictions to a discrete-event simulation model we developed previously [15].

We are not the first to suggest using MDPs to guide decision-making in RF systems. Zhao et al. [24] propose the use of MDPs for guiding what they call *opportunistic s pectrum acces s* (the ability of secondary users to identify and exploit instantaneous spectrum opportunities that arise because of the bursty traffic patterns of primary users). Tradeoffs between optimality and complexity in such cases are examined by Djonin et al. [7]. Akbar and Tranter [2] use hidden Markov models (HMMs) to model and predict the spectrum occupancy of licensed radio bands for this same purpose.

Markovian models have been used to characterize other properties of RF systems as well. Wang et al. [22] use a Markov transition system to characterize different handoff delays associated with connections in cognitive radio networks. Geirhofer et al. [10] propose a continuous-time semi-Markov model of a WLAN's behavior, towards a better understanding of primary users' activities.

3. USING MARKOV DECISION PROCESSES FOR RF SPECTRUM MANAGEMENT

The management of RF spectrum requires a number of allocation decisions (e.g., admission, placement, overlap, and modulation) to be made. If a Markov Decision Process (MDP) model can faithfully represent the system being modeled, one then can represent these decisions as actions within the MDP and generate a value-optimal policy, with a prescribed action corresponding to each reachable state of the system. In this work, we concentrate on admission, placement, and overlap decisions, i.e., whether or not a message should be allowed to use spectrum resources and if so which channel. As future work, we also anticipate that other decisions (e.g., modulation choices) could be modeled effectively in a similar manner.

We start by describing the system model that we use to represent the RF spectrum being managed. This is followed by a brief introduction to MDP models, and then a description of the specific MDP we use for admission control.

3.1 RF Spectrum System Model

In our system model, which follows the one we developed previously [15], the system has a block of radio spectrum, divided into some number of channels, and a central controller that makes decisions about the allocation of those channels. Messages arrive via a Poisson process, and are allocated to a channel if admitted, and depart the system if not admitted. For the purposes of this evaluation, we limit the decision-making to admission and placement (including whether or not messages are allowed to overlap). The mean message arrival rate is denoted by λ, and message sizes (i.e., durations) are assumed to be uniformly distributed with mean 1/μ (following the convention in queuing theory that μ is a service *rate*). The total rate of departure at any specific time, therefore, is proportional to the number of messages in the system. Multiple messages can be allocated into one channel (i.e., they can *overlap*); in that case, the success of a message delivery is a function of the RF channel model, which we describe next.

The RF channel model characterizes the success (or failure) of message delivery in terms of environmental factors and conflicts due to common channel occupancy [12]. The environmental factors (e.g., background noise, reflections, etc.) are aggregated into a single term, denoted P_{env}, representing the probability of a message delivery failure due to these factors. Message conflict is similarly characterized by a single term, denoted P_{conf}, which parameterizes a Bernoulli model of message failure. The probability that an individual message is successfully delivered, denoted P_{succ}, is therefore

$$P_{succ} = (1 - P_{env})(1 - P_{conf})^{N_m - 1}$$

where N_m is the number of messages sharing the channel.

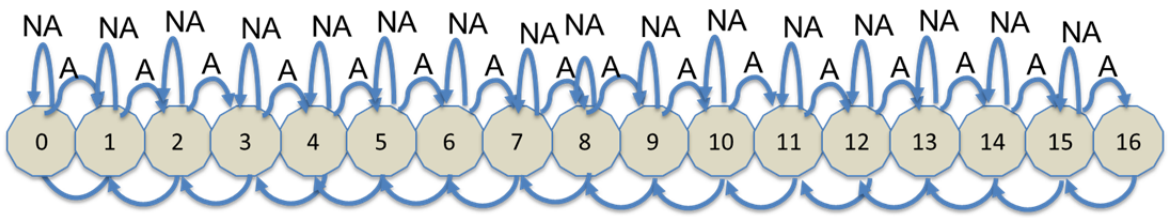

Figure 1. 4 channel, up to 16 message, Markov decision process state transition diagram.

3.2 Background on the MDP Formalism

The following description is derived from Tidwell et al. [21], which applies an MDP model to processor allocation problems[1]. An MDP is a five-tuple (X, A, P, R, γ) with *states* X and *actions* A, a *transition system* P with probabilities $P(y|x,a)$ of transitioning from state x to y on action a, and a *reward func tion* R that specifies the immediate value of each action in each state of the system. The *discount factor* $\gamma \in [0,1)$ defines how potential future rewards are weighed against immediate rewards when evaluating the impact of taking action a in a given state. A policy π for an MDP maps states in X to actions in A. At each discrete decision epoch k the agent observes the state of the MDP x_k, then selects an action $a_k = \pi(x_k)$. The MDP then transitions to state x_{k+1} with probability $P(x_{k+1}|x_k, a_k)$ and the controller receives reward $r_k = R(x_k, a_k, x_{k+1})$. Given the discount factor γ, the value of a policy, denoted V^π, is the expected sum of long-term, discounted rewards obtained while following that policy:

$$V^\pi(x) = E\left\{\sum_{k=0}^{\infty} \gamma^k r_k \,\middle|\, x_0 = x, a_k = \pi(x_k)\right\}.$$

$$R^\pi(x) = \sum_{y \in X} P(y|x, \pi(x)) V^\pi(y)$$

denotes the expected reward obtained when executing action $a = \pi(x)$ in state x. Then we may equivalently define V^π as the solution to the linear system

$$V^\pi(x) = R^\pi(x) + \gamma \sum_{y \in X} P(y|x, \pi(x)) V^\pi(y)$$

for each state x. When $|R^\pi(x)|$ is bounded for all states, the discount factor γ prevents V^π from diverging for any choice of policy, and can be interpreted as the prior probability that the system persists from one decision epoch to the next [13]. In practice this value is almost always set very close to 1 and in this work we set γ to 0.99.

There are several techniques for computing the value-optimal policy for an MDP with finite state and action spaces [16]. These techniques calculate the optimal action, $\pi^*(x)$, for every state $x \in X$:

$$\pi^*(x) = \underset{a \in A}{\operatorname{argmax}} \left\{ R(x,a) + \gamma \sum_{y \in X} P(y|x,a) V^*(y) \right\}$$

where the optimal value, $V^*(x)$, is given by:

$$V^*(x) = \max_{a \in A} \left\{ R(x,a) + \gamma \sum_{y \in X} P(y|x,a) V^*(y) \right\}.$$

The value-optimal policy is the policy that optimizes long term value, in contrast to immediate reward.

3.3 MDP Use for Admission Control

Figure 1 shows the structure of the continuous-time MDP we propose to use for admission control. Each state $x = i$ encodes the number of messages occupying a channel (i.e., currently being transmitted). Our MDP model assumes a perfect allocation of messages to channels, so if there are C channels and $x \leq C$, each message is assumed to be allocated to a distinct channel. If $x > C$, we assume the number of potential conflicts (i.e., messages transmitting on a common channel) is the minimum possible.

For admission control, the two possible actions are to *allocate* or *not allocate*. One of these actions is taken whenever a message arrives in the system. In Figure 1, actions to allocate are indicated by edges labeled A in the transition from a state i to the neighboring state $i+1$. Similarly, actions to not allocate are indicated by edges labeled NA and are self-loops (i.e., transition back into the same state rather than to a new state).

Since the Poisson arrival process has mean rate λ, the transition rates for the edges labeled A and NA are both λ. The MDP treats the set of messages as a traditional birth-death process, so the departure rate is determined by the mean service rate to be $x\mu$.

The above described (continuous-time) model is converted into a discrete-time MDP by adding self-loops and converting the transition rates to transition probabilities using the uniformization technique described by Grassman [11] with uniform rate parameter δ, which is set to be greater than the largest rate in the continuous-time model[2]. As a result, the probability of an arrival is λ/δ (for an action to accept) and the probability of a departure is $x\mu/\delta$. The probability of a self-loop is $1-(\lambda+x\mu)/\delta$ (for an action to accept) and $1-x\mu/\delta$ (for an action to not accept).

Value (from the reward function) is accrued on departure transitions, in an amount equal to the expected duration of the message time (equal to its size) multiplied by the probability of a successful transmission, P_{succ}.

Informally, we expect value-optimal admission decisions made under these conditions to result in an admission policy of accepting incoming messages up to some system occupancy and

[1] Despite the differing resources (spectrum vs. processor cycles) and semantic models (Bernoulli vs. time-utility scheduling) involved, our formulation of the MDP in this work is similarly motivated by the challenge of optimal resource use considered in that work.

[2] In the literature the uniform rate parameter is frequently represented by γ, but that symbol is already in use as part of the MDP definition.

then not accepting new messages at any higher occupancy, with the acceptance threshold being influenced in part by the value of P_{conf}, which parameterizes the penalty of sharing individual channels. This policy is known to be optimal for real-world spectrum allocation circumstances, e.g., when using FM modulation.

This intuition is confirmed by setting the value of P_{conf} to 0.97 in a 4-channel system and solving for the value-optimal policy using the MDP. At this high value of P_{conf} (i.e., high probability of message delivery failure due to conflict), sharing of channels is unlikely to benefit throughput, and the policy that is chosen via the MDP is to allocate up to 4 messages, but no more.

We expand this investigation to $P_{conf} = 0.7$ (i.e., conflicts have a lower probability of causing message delivery failure) in the same 4-channel system and again solve for the value-optimal policy. In this case, the resulting policy is to allocate up to 8 messages, but no more. This time each channel is potentially shared by up to 2 messages.

The intuition associated with these experimental results is that as P_{conf} gets smaller, the likelihood of overlapping messages being successfully delivered increases (i.e., P_{succ} goes up). Therefore, a policy with greater allowed message overlap is value-optimal.

With confirmation that our MDP is making reasonable admission decisions, we next proceed to assess the appropriateness of some of the modeling assumptions made during the development of the MDP model.

4. EVALUATION

In this section we evaluate the use of a Markov decision processes for RF spectrum allocation by empirically comparing decisions and performance predictions made by our MDP model with performance predictions made by a pre-existing discrete-event simulation (DES) model [15].

4.1 Approach

A key premise of this paper is that there is inherent value in using MDP models to manage and control RF spectrum resources. Our approach to evaluate that premise is to compare our MDP model with a different (i.e., DES) model. To appreciate both the benefits and limitations of this approach, it is important to distinguish what this empirical comparison does and does not validate. Specifically, any modeling assumption that is made in common by both the MDP model and the DES model will not be tested by this approach. We first articulate some such assumptions:

1. The arrival process of messages is assumed to be Poisson with a given mean arrival rate.
2. Messages each consume one channel of RF spectrum (i.e., the spectrum is decomposed into discrete, equal-sized channels) and collectively have a given mean duration.
3. The underlying channel model that predicts the success or failure of an individual message delivery, based on environmental factors and/or conflict with other messages, is common across the MDP and DES models.
4. Messages that fail depart the system. We do not model retries.

Although one certainly may question these assumptions, it is impractical to test them given the available experimental infrastructure, since the DES model (which also makes those same assumptions) is our comparison vehicle. As we describe in Section 3, these assumptions also come directly from the RF spectrum system model we consider.

While the above list describes what we will not test, the purpose of our evaluation is to assess the three primary hypotheses (stated in Section 1), which we *can* test:

1. the independence of message throughput on the distribution of message durations,
2. the insignificance of imperfect allocations, and
3. that value-optimal policies chosen by the MDP are at least locally optimal according to the DES model.

4.2 Discrete-Event Simulation Model

The discrete-event simulator maintains an explicit representation of the set of channels within the RF spectrum range being modeled, including occupancy of each channel over time by specifically which set of messages. Many features of the discrete event simulator are unnecessary for this evaluation (e.g., it supports a binary buddy algorithm for spectrum allocation) but since the present investigation is constrained relative to the simulator's capabilities (e.g., our comparison is limited to messages that occupy only a single channel, therefore allowing for a simple greedy allocation algorithm) we will only describe features that are directly relevant to our evaluations.

The DES uses traditional event-driven simulation techniques, in which state changes to the modeled system are represented by time-stamped events that are maintained in a time-ordered priority queue. The event with the smallest time-stamp is removed from the queue, the state change represented by that event is executed, and any subsequent future state changes implied by the event's execution are scheduled in the priority queue.

The events supported by the simulator include *message arrival* and *message departure*. As part of the execution of a message arrival event, the simulator performs a greedy allocation algorithm (i.e., it allocates the newly arrived message to a channel with the fewest conflicts with other, pre-existing messages currently using the spectrum). During the execution of a message departure event, the Bernoulli model of Section 3.1 is used to determine whether or not the message was successfully delivered.

In [15], performance predictions made by the simulator are compared with an *M/U/c/c* queueing model (i.e., *M*arkovian arrival process, *U*niformly distributed service process, *c* servers, and *c* total jobs allowed in the system). To support the evaluations we wish to perform in this work, the simulator was extended it two specific ways. First, the uniform distribution assumption for message duration was expanded to also include the option for an exponential distribution. Second, the effectiveness of the greedy allocation algorithm was measured by counting the number of messages that were delivered under imperfect allocation decisions (i.e., the message was delivered using a shared channel when a free channel was available but, at the time, unknown to the simulator).

4.3 Experimental Predictions

We now describe experimental predictions related to each hypothesis. Each experiment is designed to assess the distinctions between the MDP model and the DES model. The first two hypotheses are assessed by experiments utilizing the discrete-event simulator. The third hypothesis is assessed initially by comparing the throughput performance predictions of the MDP

model to those of the DES model, followed by using the simulator to examine local optimality of the decisions made by the MDP.

The first experiment evaluates the impact of the distribution of message durations on the predicted throughput of the system. The MDP model makes the standard memory-less assumption, modeling message durations via an exponential distribution. The DES model follows a common convention in RF systems and models message durations via a uniform distribution. To test our first hypothesis we modify the simulator to support an exponential distribution for message durations, and compare the throughput predictions for uniformly distributed message durations with that for exponentially distributed message durations. We intentionally perform this experiment exclusively in the simulation model. Our experimental prediction is that there will not be a significant difference between the throughputs for the two (different) distributions of message duration. In effect, we are testing whether or not the *insensitivity property* that is well established for Erlang-loss systems [17] holds here.

The second experiment evaluates the impact of imperfect allocations on the throughput of the system. The MDP model implicitly assumes that allocation decisions are perfect (i.e., if a free channel exists in the spectrum, some other channel isn't shared). The DES model makes no such assumption, but rather implements the specific actions of a greedy allocator. By measuring the frequency of imperfect allocations (which do not meet the ideal assumptions made by the MDP model) in the simulation, we can assess the impact of this assumption on the performance predictions made by the MDP model. We predict that the frequency of these imperfect allocations is sufficiently small that its effect on throughput is within the statistical variation of the throughput predictions made by the simulator. As in the previous experiment, we use the discrete-event simulator to assess the validity of the perfect channel allocation assumption.

The third experiment is designed to assess the appropriateness of the MDP's value-optimal policy decisions. We predict that a reasonable correspondence (based on local optimality) exists between the performance predictions of the proposed MDP model and the DES model, such that value-optimal policy decisions made by the MDP correspond to optimal throughput predictions by the DES model. As true optimality is computationally impractical to test, we explicitly check for local optimality (i.e., the policy chosen by the MDP is at least a locally optimal choice as predicted by the DES model). Starting from the value-optimal policy chosen by the MDP, we ask the DES to predict performance given that policy and several "nearest neighbor" policies, with the neighborhood chosen to represent what we mean by locality.

5. EXPERIMENTAL RESULTS

In this section we describe the experiments we conducted to evaluate the hypotheses discussed above. Section 5.1 describes design details and infrastructure used to conduct our experiments, and Section 5.2 presents and discusses their results.

5.1 Experimental Setup

The MDP and DES models are configured to evaluate the models' behaviors with either a single channel or four channels of spectrum available for allocation. Each of these channels can have up to four levels of redundant use (or reuse). We describe the set up for each of these experiments, and how the evaluations take place. The MDP model is configured using a Bernoulli reward function (described in Section 3.1)[3]. All throughput predictions are made over a range of input rates that provide normalized offered load, ρ, between 0.5 and 2.5 (i.e., $0.5 \leq \rho = \lambda/\mu \leq 2.5$), P_{conf} ranging from 0.1 to 0.9, and $P_{env} = 0$ (the latter since P_{env} was shown in [15] to have a simple linear impact on throughput). All DES model throughput results are analyzed using the method of batch means, with 100 independent runs decomposed into 10 batches of size 10.

5.1.1 DES experiments to assess significance of message duration distribution

Our MDP model assumes exponentially distributed message durations. The DES model assumes uniformly distributed message durations. We configure the DES to test both uniform and exponential message durations and evaluate the impact. We plot the predicted throughput for both uniform and exponential message durations (using single standard deviation whiskers) for a variety of channel counts and system utilization. We call this experiment *distribution assessment*.

5.1.2 DES experiments to count and assess significance of imperfect allocations

The DES also was modified to count imperfect allocations. These were compared to the normal allocations, constructing a ratio for comparison. We modified the simulator to record the number of imperfect spectrum allocations by counting whether or not a free channel is available each time a message completes using a shared channel.

The throughput ratio (R_T) provides a normalized measure of variability in throughput that can be used to judge the impact of imperfect allocations. Defined as twice the coefficient of variation, its intent is to provide a comparison point for imperfect allocations (i.e., if the ratio of imperfect allocations to total allocations is lower than R_T, they can be considered to be infrequent enough to be within the normal stochastic variation inherent in the simulation model).

$$R_T = \frac{2 \times \textit{throughput std. dev.}}{\textit{mean throughput}}$$

The ratio of total messages transmitted relative to the number of imperfect allocations provides a measure of their significance.

$$R_I = \frac{\textit{imperfect allocations}}{\textit{total allocations}}$$

We contrast the two ratios (R_T and R_I) to evaluate the effects of imperfect allocations on the two models' throughput predictions. We call this experiment *imperfect allocations*.

5.1.3 Comparison of MDP to DES

We plot throughput predictions made by the MDP and DES models for different experimental configurations and compare trends between the two models' results. We then configure the MDP to generate a value-optimal policy for spectrum allocation. We test the local optimality of the allocations dictated by that policy by evaluating DES predicted throughput in configurations in the neighborhood of those that use the value-optimal policy. In

[3] We also considered a simple reward function based on the number of messages allocated, without considering message loss. Since the Bernoulli reward function more accurately describes semantics of FM modulation, for brevity we omit discussion of other reward functions in the rest of this paper.

these experiments, the neighborhood is defined by varying the allowed message overlap. We call these experiments *throughput evaluation*.

5.2 Experimental Results

Here, we are interested in exploring the quantitative results of the set of experiments described above. We organize the results according to the three hypotheses that we wish to evaluate:

5.2.1 Distribution Assessment Experiment

The intent of the distribution assessment experiment is to determine whether or not the mean of the message size distribution is sufficient to characterize the achievable throughput (i.e., how important is the shape of the distribution).

Figure 2 plots message throughput predicted by the DES model as a function of offered load for a pair of different circumstances:

1. $P_{conf} = 0.1$, single channel (C = 1), no overlap of messages (L = 1), and offered load ranging from 0.5 to 2.5.
2. $P_{conf} = 0.1$, four channels (C = 4), overlap of up to 4 messages (L = 4), and offered load ranging from 0.5 to 2.5.

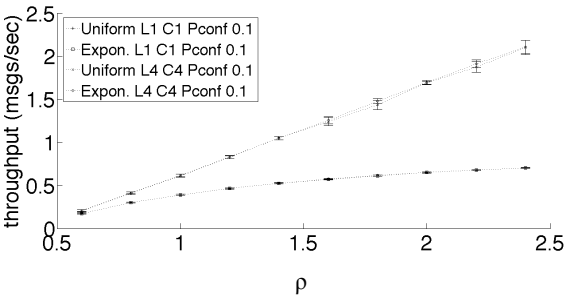

Figure 2. Throughput vs. offered load for both uniform and exponential distribution of message durations, ρ is offered load (λ/μ). Whiskers represent standard deviation.

For each of these circumstances, the simulator is configured to use both the original DES model assumption (uniformly distributed message duration) and the MDP model assumption (exponentially distributed message duration). Points represent the mean over 100 simulation executions and the whiskers represent standard deviation of 10 batches (each of size 10).

First, we observe that the throughput predictions are reasonable, with greater throughput achievable with greater capacity. As is readily apparent in the plots, the distinction in throughput between these two distribution assumptions is quite small and is well within the expected deviations due to statistical variation. This evidence thus strongly supports hypothesis 1, that "performance can be reasonably characterized by the mean of message durations and is relatively insensitive to their distribution."

To be clear, we have only truly verified the correspondence between the uniform distribution (commonly used in the RF literature and the original model assumption present in the discrete-event simulation model) and the exponential distribution (used in the Markov decision process model). Where this hypothesis might yet not be true is for extreme values of the true distribution's variance. E.g., deterministic (fixed duration) messages with zero variance or more heavy-tailed distributions with large variance. We leave this investigation for future work.

5.2.2 Imperfect Allocations Experiment

Hypothesis 2 posits insignificant performance variation due to imperfect allocations that are abstracted away in the MDP model. The imperfect allocations are recorded using the simulator to assess the significance of ignoring these occurrences. We record the occurrences when a channel is empty but allocations are yet made on already allocated channels. We examine:

1. The four channel system (C = 4), overlap of up to 4 messages, and offered load ranging from 0.5 to 2.5, with
2. $P_{conf} = 0.1$, 0.5, and 0.9.

We constrain the experiment to the four channel system with overlap since the other configurations cannot exhibit imperfect allocations. The 1000 independent DES simulations executed for each value of P_{conf} constitute over 500,000 allocations. The results are presented in Table 1. Recall that R_I is the ratio of imperfect allocations to total allocations and that R_T is twice the coefficient of variability in the resultant throughput predicted by the simulator.

Table 1. Frequency of imperfect allocations.

P_{conf}	R_I	R_T
0.1	0.0160	0.0156
0.5	0.0165	0.0156
0.9	0.0162	0.0156

These results indicate that only a limited number of allocations are affected by the imperfect allocations not accounted for with the MDP model (approximately equal to 2 standard deviations of the simulation output's statistical variability). As a result, the maximum impact on throughput is well within the statistical variation illustrated in Figure 2, and this evidence supports hypothesis 2. Also, the low impact of imperfect allocations implies that the greedy allocation algorithm is working quite well, i.e., essentially indistinguishable from a perfect allocator.

5.2.3 Throughput Evaluation Experiments

Hypothesis 3 examines the use of the MDP model for evaluating optimal throughput characterizations. We assess this hypothesis by examining two things: (1) throughput predictions made by both the DES and MDP models; and (2) the local optimality (as confirmed by the DES model) of value-optimal policies chosen by the MDP. The throughput predictions are from the following set of experiments:

1. $P_{conf} = 0.9$, using a single channel (C = 1), no overlap of messages (L = 1), and offered load ranging from 0.5 to 2.5 (Figure 3).
2. $P_{conf} = 0.5$, using four channels (C = 4), no overlap of messages (L = 1), and offered load ranging from 0.5 to 2.5 (Figure 4).
3. $P_{conf} = 0.1$, using four channels (C = 4), overlap of up to 4 messages (L = 4), and offered load ranging from 0.5 to 2.5 (Figure 5).

The first experiment is limited to no overlap of messages primarily because with such a large value of P_{conf}, overlapping messages do not effectively improve throughput in any event.

With only a single channel and no message overlap allowed, we expect a maximum throughput bounded above by an individual channel's capacity. The latter two experiments explore the use of additional channels and message overlap for two different values of P_{conf}. Here, we would expect to see some variation in achievable throughput between the two experiments. In each of the above experiments, we are comparing the throughput predictions of the MDP to that of the DES. To accomplish this while manually controlling the MDP policy (i.e., number of overlap levels allowed), we disable the value-optimal policy evaluation within the MDP and manually set the policy we wish to explore. This manual policy setting action transforms the Markov decision process into a traditional Markov process, for which we can determine the throughput by solving for the steady-state occupancy probabilities for each state *x* in the original MDP.

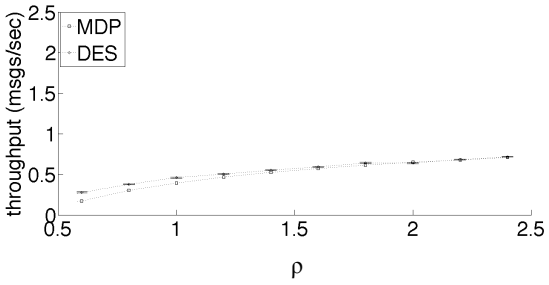

Figure 3. MDP vs. DES model throughput predictions for $P_{conf} = 0.9, C = 1, L = 1$.

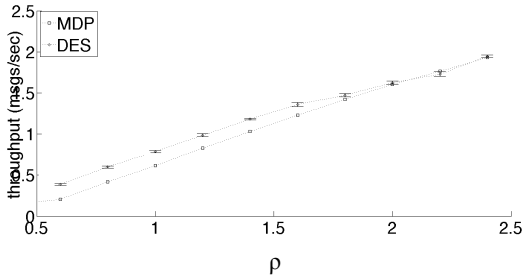

Figure 4. MDP vs. DES model throughput predictions for $P_{conf} = 0.5, C = 4, L = 1$.

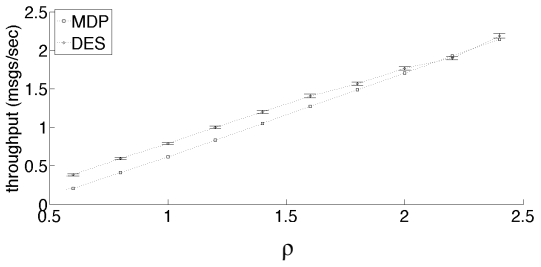

Figure 5. MDP vs. DES model throughput predictions for $P_{conf} = 0.1, C = 4, L = 4$.

As was expected, the maximum throughput achievable with only a single channel is quite limited. Achievable throughputs, however, increase as additional resources are made available.

All three plots show reasonable agreement between the throughput predictions made by the MDP model and the DES model. While the MDP model's throughput predictions are not always within one standard deviation of the DES model's mean throughput, even when they separate it is not by far. Additionally, as the offered load increases, the separation between the two models actually diminishes. Given that imperfect information is less important to an admission algorithm at low load, we are much more interested in the correspondence between the two models under high-load conditions.

After concluding that we are able to model the throughput using the MDP model, we re-enable the ability of the MDP to choose a value-optimal policy. This allows us to test the (local) optimality of the policy chosen by the MDP by using the simulator. We confirm local optimality by doing a local neighborhood search with the discrete-event simulator and assessing whether or not the policy chosen by the MDP corresponds to the local throughput maximum.

The experiment is run with $P_{conf} = 0.7$, which gives a value-optimal admission policy (provided by the MDP) of using up to 2 levels of overlap (i.e., up to 2 messages per channel, but no more). Using the simulator, we assess the throughput predictions for $L = 1$ (no overlap), $L = 2$ (value-optimal according to the MDP), and $L = 3$. The mean throughput predictions from the DES model are shown in Table 2.

Table 2. Local Optimality.

L	1	2	3
throughput	1.93	2.11	2.07

Although the separation between these throughput predictions is not very large, the throughput for $L = 2$ is clearly above that of $L = 1$ and $L = 3$. The value-optimal policy chosen by the MDP is confirmed to be locally optimal as assessed by the DES model. This provides evidence for confirming hypothesis 3.

6. CONCLUSIONS AND FUTURE WORK

This paper has assessed the use of MDP models for making management decisions for RF spectrum. We have formulated 3 distinct hypotheses, developed and conducted experiments to assess each of these hypotheses, and the empirical results all support the confirmation of the hypotheses. We therefore conclude that the use of MDPs for RF spectrum management is a reasonable (and potentially fruitful) path to explore.

The example MDP we present in this paper is suitable for admission decisions, and we have provided evidence that (for this simple case) a greedy placement algorithm is effective. In the future, we plan to expand the MDP model to incorporate additional allocation issues (e.g., explicitly including choice of modulation) and more robust channel models (e.g., to include modulation impact and/or distance measures).

7. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grants CCF-0448562 and CNS-0931693, and by Object Computing, Inc.

8. REFERENCES

[1] Abundo, M., Cardellini, V., and Presti, F. L. 2011. An MDP-based Admission Control for a QoS-aware Service-oriented System. In *Proc. of IEEE 19th Int'l Workshop on Quality of Service.* (June 2011), 1-3.

[2] Akbar, I. A. and Tranter, W. H. 2007. Dynamic Spectrum Allocation in Cognitive Radio using Hidden Markov Models:

Poisson Distributed Case. In *Proc. of IEEE SoutheastCon*. (March 2007), 196-201.

[3] Anchora, L., Mezzavilla, M., Badia, L., and Zorzi. M. 2011. Simulation models for the performance evaluation of spectrum sharing techniques in OFDMA networks. In *Proc. of 14th ACM Int'l Conf. on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. (2011), 249-256.

[4] Beard, J. C. and Chamberlain, R. D. 2013. Use of Simple Analytic Performance Models for Streaming Data Applications Deployed on Diverse Architectures. In *Proc. of IEEE Int'l Symp. on Performance Analysis of Systems and Software*. (April 2013), 138-139.

[5] Bethke, B., Bertuccelli, L. F., How, J. P. 2008. Experimental Demonstration of Adaptive MDP-Based Planning with Model Uncertainty. In *Proc. AIAA Guidance, Navigation and Control Conf.* (Aug. 2008).

[6] Dillinger, M., Madani, K., and Alonistioti, N. 2003. *Software Defined Radio: Architectures, Systems, and Functions.* Wiley.

[7] Djonin, D. V., Zhao, Q., and V. Krishnamurthy. 2007. Optimality and Complexity of Opportunistic Spectrum Access: A Truncated Markov Decision Process Formulation. In *Proc. of IEEE Int'l Conf. on Communications.* (June 2007), 5787-5792.

[8] Feeney, L. M. 2012. Towards trustworthy simulation of wireless MAC/PHY layers: a comparison framework. In *Proc. of 15th ACM Int'l Conf. on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. (2012), 295-304.

[9] Fette, B. 2006. *Cognitive Radio Technology*. Elsevier Science & Technology Books.

[10] Geirhofer, S., Tong, L., and Sadler, B. M. 2006. A Measurement-Based Model for Dynamic Spectrum Access in WLAN Channels. In *Proc. of IEEE Military Communications Conf.* (Oct. 2006), 1-7.

[11] Grassman, W. K. 1977. Transient Solutions in Markovian Queueing Systems. *Computers & Operations Research.* 4(1):47-53 (1977).

[12] Hou, I.-H., Borkar, V., and Kumar, P.R. 2009. A Theory of QoS for Wireless. In *Proc. of IEEE Int'l Conf. on Computer Communications.* (April 2009), 486-494.

[13] Kaelbling, L. P., Littman, M., and Moore, A. 1996. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*. 4:237-285 (1996).

[14] Lee, B. C. and Brooks, D. 2008. Roughness of Microarchitectural Design Topologies and Its Implications for Optimization. In *Proc. of IEEE Int'l Symp. on High Performance Computer Architecture*. (2008), 240-251.

[15] Meier, J., Gill, C., and Chamberlain, R. D. 2011. Towards More Effective Spectrum Use Based on Memory Allocation Models. In *Proc. of 35th IEEE Computer Software and Applications Conf.* (July 2011), 426-435.

[16] Puterman, M. L. 1994. *Markov Decision Processes*. Wiley.

[17] Sevastyanov, B. A. 1957. An Ergodic Theorem for Markov Processes and Its Application to Telephone Systems with Refusals. *Theor. Probability Appl*. 2:104-112 (1957).

[18] Shani, G., Brafman, R. I., and Heckerman, D. 2002. An MDP-based Recommender System. In *Proc. of Conf. on Uncertainty in Artificial Intelligence.* (2002), 453-460.

[19] Staple, G. and Werbach, K. 2004. The End of Spectrum Scarcity. *IEEE Spectrum*. 41(3):48-52 (Mar. 2004).

[20] Szabó, B. and Babuška, I. 1991. *Finite Element Analysis*. Wiley.

[21] Tidwell, T., Glaubius, R., Gill, C, and Smart, W. 2010. Optimizing Expected Time Utility in Cyber-Physical Systems Schedulers. In *Proc. of 31st Real-Time Systems Symposium*. (Nov. 2010).

[22] Wang, C.-W., Wang, L. C., and Adachi, F. 2010. Modeling and Analysis for Reactive-Decision Spectrum Handoff in Cognitive Radio Networks. In *Proc. of Global Telecommunications Conf.* (Dec. 2010), 1-6.

[23] Zhao, Q. and Sadler, B. M. 2007. A Survey of Dynamic Spectrum Access. *IEEE Signal Processing Magazine*. 24(3):79-89 (May 2007).

[24] Zhao, Q., Tong, L., Swami, A., and Chen, Y. 2007. Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework. *IEEE J. on Selected Areas in Communications*. 25(3):589-600 (April 2007).