ECAI 2012
Luc De Raedt et al. (Eds.)

doi:10.3233/978-1-61499-098-7-690

Coordinated Exploration with a Shared Goal in Costly Environments

Igor Rochlin and **David Sarne**[1] and **Moshe Laifenfeld**[2]

Abstract.

The paper studies distributed cooperative multi-agent exploration methods in settings where the overall benefit of an opportunity is the minimum of individual findings and the exploration is costly. The primary motivation for the model is the multi-channel cooperative sensing problem which draws from the inter-vehicular cognitive offload paradigm. Here, vehicles try to coordinate an offload channel through a dedicated common control channel, and the resulting quality of the channel eventually selected is constrained by the individual qualities. Similar settings may arise in other multi-agent settings where the exploration needs to be coordinated. The goal in such problems concerns the optimization of the process as a whole, considering the tradeoff between the quality of the solution obtained for the shared goal and the cost associated with the exploration and coordination process. The methods considered in this paper make use of parallel and sequential exploration. The first approach is more latency-efficient, and the latter is shown to be more cost-effective. The strategy structure in both schemes is threshold-based, and the thresholds which are analytically derived in this paper can be calculated offline, resulting in a very low online computational load. A comparative illustration of the methods' performance is given using a synthetic environment, emphasizing the cost-latency tradeoff.

1 Introduction

In many multi-agent settings agents need to individually engage in exploration of the opportunities available to them in the environment in which they are operating in order to pick one that will satisfy a desired goal [2]. The purpose of the exploration is to reason about the nature and value of the different opportunities whenever such information is a priori unknown. Since such exploration is inherently costly (either involves direct monetary costs or the consumption of some of the agent's resources) the goal of the agent is not necessarily to find the opportunity associated with the maximum value, but rather to maximize the overall benefit, defined as the value of the opportunity eventually picked minus the costs accumulated during the exploration process [8].

This kind of exploration process becomes more complex whenever conducted cooperatively by several agents. For example: when the agents are robots that need to evaluate several potential locations for mining a certain mineral on the face of Mars (if all robots need to cooperate in the mining process then the goal is to find the location to which the latest time it takes any of them to arrive is minimized) [10]; a group of buyers that need to evaluate several potential sellers for buying different products [20]; when individuals need to decide on a restaurant to dine in as a group (where the group's benefit is the minimum of the individual valuations); and secondary users in Dynamic Spectrum Access applications that need to evaluate different frequencies in order to establish a common communication link [17]. In such cases, the agents not only need to consider other agents' performance, but also the overhead associated with the coordination between them along the process.

In this paper, we formally introduce a model of a multi-agent exploration in which: (a) each agent sees a different individual value in each available opportunity; (b) the value of the opportunity picked eventually (in terms of goal satisfaction or overall quality) is a function of the individual values the different agents see in that opportunity; and (c) the performance measure also takes into consideration the costs accumulated by the different agents along the process. There are many settings characterized by such a cooperative exploration model. For example, consider the vehicular ad hoc networks (VANETs) domain where cooperative sensing of Secondary Users (SUs) takes place. Here SUs (vehicles) interested in sharing high-capacity non-safety related data need to find appropriate shared spectrum bands (channels) using Cognitive Radio (CR) technology [4, 22]. The quality (the equivalent to opportunity value in this case) of the band eventually used is the minimum quality experienced by the different SUs, and in order for a SU to measure its perceived quality of a given band it needs to execute a costly sensing. The coordination between the different SUs is done through a dedicated short-range communication channel (DSRC), which is a rather limited resource [4]. Alternatively, consider several agents, each representing a participant in a planned trip. All participants need to get to the same starting point from which they will all share transportation to the destination, and participants differ in the time it takes them to get to different potential starting points. The agents need to cooperatively explore the benefit from starting at the different starting points, sharing their arrival time valuations to the different starting points. The value of the starting point eventually picked is the maximum time it takes any of the participants to get to it, as the shared transport device will be able to leave only when all participants arrive.

To the best of our knowledge, this is the first attempt to analyze the above model in full. As discussed in the related work section, former cooperative exploration literature usually does not consider the overall quality to be a function of the individual qualities of the different agents or does not adequately take into consideration the coordination costs. The analysis presented in this paper considers two main cooperative distributed exploration schemes. In the first, the exploration process is executed in parallel, which is favorable when latency is of importance. The second is based on sequential exploration, which takes longer to execute, however is more efficient

[1] Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel
[2] General Motors, Advanced Technical Center - Israel

in terms of the overall expected cost [15]. The paper provides the expected-benefit maximizing exploration strategy for each of the two schemes. Furthermore, since both mechanisms are threshold-based, the coordination cost can be substantially reduced simply by the use of an intelligent protocol that is tightly coupled with the exploration strategy.

2 Related Work

While multi-agent cooperation and coordination mechanisms have been widely studied in AI over the years, most of such work was dedicated to the way such cooperation is formed and consequently concerned issues such as the optimal division of agents into disjoint exhaustive teams [11, 19, 23], division of the payoffs resulting from cooperation [23] and enforcement methods for interaction protocols [13]. Very few works considered the cooperative execution of a task once the cooperation is formed [9].

The problem of a single agent engaged in costly exploration of the type discussed in this paper has been extensively researched in literature [8, 12]. While the analysis given in the literature adequately analyzed the tradeoff between the potential gains and the resulting costs from extending the exploration process, the transition to a multi-agent exploration is not trivial as these do not take into consideration coordination aspects (and indeed the nature of the solution is different than the one presented in this paper). Those few works that do go beyond a single agent's costly exploration considerations merely consider settings where the agents do not necessarily need to commit to the same opportunity [5, 7, 17, 20] or where the value of each opportunity to all agents is common thus, it is sufficient that only one of the agents explore any specific opportunity [10]. These inherent differences typically enable solving the exploration problem through its mapping to a single agent's exploration problem and the coordination between the different agents significantly degenerates. To the best of our knowledge, the problem of a group of agents that needs to cooperatively evaluate opportunities and choose one while the benefit of each depends on its values as seen by others, while facing coordination costs, has not been investigated to date.

Finally, we note that the concept of cooperative exploration is a common motive in Dynamic Spectrum Access and Cognitive Radio applications [16, 18, 21], due to their inherent need for spectrum sharing. Nevertheless, the rich literature of this domain usually focuses on spectrum optimization (considering a global view of all channels, aiming to improve channels utilization) [1, 3, 6] rather than considering the individual agents' problem of executing an ad-hoc exploration for a single channel. As such, cooperative exploration is applied to usages such as improving the estimations regarding channels qualities (e.g., have several terminals sense a given channel in order to better estimate its availability [16]), sharing the load of collecting observations [21] and efficiently collecting information that can then be transferred to a fusion center, which performs a joint optimization of the different sensing measurements for all channels [18].

3 The Model

We consider a setting where k individual cooperative agents need to engage in exploration of possible available opportunities, all satisfying a well-defined shared goal. Each agent may see a different value in each opportunity and the values are a priori unknown. The values are assumed to be drawn from a probability distribution function, denoted $f(y)$, with which all agents are acquainted. In order to obtain its actual value from a specific opportunity the agent needs to explore it, incurring a non-negligible cost (e.g., monetary or in terms of resources it needs to consume), denoted c_e. The cost c_e is common to all agents as they are all assumed to share the same exploration technology. The agents need to, eventually, cooperatively choose a single opportunity that will be used to satisfy their shared goal. The model assumes that the agents can explore as many opportunities as they require and that the overall value from the specific opportunity eventually picked is a function of its individual values as perceived by the different agents. Specifically, the analysis given in the following section considers this function to be the minimum of the k values. The agents can coordinate their exploration through reliable message broadcasting (i.e., transmitting the value they see in a given opportunity through a message that is received by all other agents). The cost of broadcasting a message is denoted c_b and is common to all agents for the same technological considerations. The model assumes that both c_e and c_b are additive and can be expressed in terms of opportunity values. The goal of the agents is therefore to maximize the expected overall benefit, denoted EB, defined as the value of the selected opportunity minus the accumulated costs along the process. One important factor that needs to be considered when evaluating different cooperative exploration methods within this framework is the process latency. The model assumes that the time it takes to broadcast a message is negligible in comparison to the time it takes to explore an opportunity.

For exposition purposes, we use the Dynamic Spectrum Access application terminology in the remaining parts of the paper. Here, the agents represent secondary users that seek opportunistic channels to offload their communications and coordinate their exploration using a shared common control channel. Each secondary user measures a different quality of each channel (e.g., a channel's quality may represent its current availability probability). A channel's quality can be obtained through a sensing process, incurring a cost (e.g., handset's downtime as well as the resources (e.g., battery power) consumed as part of the operation). The secondary users need to eventually choose a single common channel which they will all use to establish their communication link. The overall quality of a channel is taken to be the minimum of the individual qualities sensed by all secondary users for that channel (e.g., in a conference call the quality of the call is constrained by the minimum quality of any of the participants' connections). The secondary users can coordinate via message broadcast through a common control channel (DSRC), which in turn also consumes some of their resources and hence is considered costly. The goal of the secondary users is to maximize the expected overall benefit, defined as the overall quality of the selected channel minus the accumulated costs along the process.

4 Analysis

The analysis considers two main cooperative distributed exploration schemes: parallel and sequential based. In both methods the channels are evaluated sequentially (in a random pre-defined order, as they are all a priori alike). The difference is in the way each channel is evaluated.

4.1 Parallel Exploration

The most straightforward coordinated exploration scheme is to have all agents simultaneously sense the same channel and broadcast their derived quality over the control channel. Then, based on the minimum sensed quality as reflected in the broadcasts, the agents decide whether to terminate the process and use that channel or to proceed and sense another channel. This exploration scheme can be fully mapped to the problem of a single agent's exploration, where the cost of evaluating each channel is $k(c_e + c_b)$ and the value obtained from it is the minimum of k values of the distribution $f(y)$. The solution to this latter problem is known to be a stationary myopic threshold-based rule [14]. According to this decision rule, the agents will terminate their exploration only when the minimum of the k quality sensings of a given channel is above a threshold r. The expected benefit from using this strategy, denoted $V(r)$ is:

$$V(r) = \frac{\int_{y=r}^{\infty} y f_{min}(y) - k(c_e + c_b)}{1 - \int_{y=-\infty}^{r} f_{min}(y)} \quad (1)$$

where $f_{min}(y)$ is the probability distribution function of the minimum of the qualities individually sensed by all agents and is given by $f_{min}(y) = kf(y)(1 - F(y))^{k-1}$ (where $F(y)$ is the appropriate cumulative density function). The expected number of channels evaluated is simply the inverse of the success probability, $\frac{1}{1-\int_{y=-\infty}^{r} f_{min}(y)}$, since this becomes a Bernoulli sampling process as channels arise independently at each iteration. Therefore, intuitively, we can interpret Equation 1 as the composition of the expected value of the minimum obtained eventually (greater than r), $\frac{\int_{y=r}^{\infty} y f_{min}(y)}{1-\int_{y=-\infty}^{r} f_{min}(y)}$, and the accumulated cost throughout the exploration $\frac{kc_e}{1-\int_{y=-\infty}^{r} f_{min}(y)} + \frac{kc_b)}{1-\int_{y=-\infty}^{r} f_{min}(y)}$ (where the first is the expected exploration cost and the second is the coordination cost).

The threshold r that maximizes the expected benefit from the process satisfies [15]:

$$k(c_e + c_b) = \int_{y=r}^{\infty} (y - r) f_{min}(y) dy \quad (2)$$

Intuitively, the value of r according to Equation 2 can be interpreted as the one with which the agents are precisely indifferent: the expected marginal benefit from continuing the exploration process and exploring an additional channel exactly equals the cost of doing so.

Single agent exploration literature which problem maps does not distinguish between the different cost components, c_e and c_b. In the multi-agent case, however, there is great importance to that fact since, by designing a better coordination, a far more efficient overall exploration process can be obtained. First, we note that since the strategy is threshold-based, the purpose of the coordination is merely to identify if all agents sensed a value greater than the threshold r. The agents can thus also settle for only broadcasting a binary signal (rather than the actual value sensed), indicating whether or not the quality found is above or below r. Furthermore, since the communication is reliable, many of the broadcasts can be saved if all agents follow a convention according to which a broadcast is made only if the sensed quality is below r. This way, the exploration will be resumed as long as at least one agent broadcasts and terminated if no broadcast received for a given channel. The process is illustrated in Figure 1(a). The expected benefit according to the improved strategy, denoted "parallel+" onwards, is:

$$V(r) = \frac{\int_{y=r}^{\infty} y f_{min}(y) dy - k(c_e + c_b F(r))}{1 - \int_{y=-\infty}^{r} f_{min}(y) dy} \quad (3)$$

The only change from (1) to (3) is the multiplication of c_b by the probability that any of the agents sense a quality lesser than r, $F(r)$.

Theorem 1. *For the "parallel+" exploration model:*
(a) The expected-benefit maximizing threshold r is given by:

$$kc_e + c_b(1 + (k-1)F(r)) = \int_{y=r}^{\infty} (y - r) f_{min}(y) dy \quad (4)$$

(b) The expected benefit when using the expected-benefit maximizing threshold r satisfies: $V(r) = r + \frac{c_b}{(1-F(r))^{k-1}}$

Proof. Equating the first derivative of (3) to zero and applying some mathematical manipulation obtains (a).
Substituting $V(r)\left(1-\int_{y=-\infty}^{r} f_{min}(y)dy\right) = \int_{y=r}^{\infty} y f_{min}(y)dy - k(c_e + c_b F(r))$ (based on (3)) in (4) obtains (b). □

The relation $V(r) = r + \frac{c_b}{(1-F(r))^{k-1}}$ is in contrast to the regular parallel model (as well as to the various other single-agent threshold-based exploration models found in literature [14]), where $V(r) = r$. While the change may seem minor, the two equations actually derive from different sets of considerations: In the new model the value of r no longer captures the indifference between resuming the exploration, obtaining $V(r)$ and terminating the exploration, obtaining r. Instead, the value of the revenue-maximizing r is derived from the threshold by which any single agent becomes indifferent to broadcasting if sensing a quality r. If broadcasting, the agent incurs a broadcast cost c_b. The marginal expected benefit $V(r) - r$ in this case is obtained, however, only if none of the other agents have sensed a quality below r (since otherwise the exploration process is resumed anyhow, and the broadcast of the agent is redundant). The probability of the latter event is $(1 - F(r))^{k-1}$, thus the indifference equation is $(1 - F(r))^{k-1}(V(r) - r) = c_b$, which transforms into $V(r) = r + \frac{c_b}{(1-F(r))^{k-1}}$.

The expected number of channels evaluated when using "parallel+" uses the same formula as in the regular parallel model, $\frac{1}{1-\int_{y=-\infty}^{r} f_{min}(y)}$ (except that the value r used is different). The expected latency in both cases equals the expected number of channels evaluated.

Finally, we note that an alternative to broadcasting only when the sensed quality is below r is to broadcast only if it is above r. In this case, however, the expected-benefit maximizing threshold is given by $kc_e + c_b(k-1)(1-F(r)) = \int_{y=r}^{\infty} (y-r) f_{min}(y)dy$, and the expected benefit of using that threshold satisfies $V(r) = r - \frac{c_b}{(1-F(r))^{k-1}}$ (as the indifference now is between potentially saving c_b, which may resume the exploration in case all other agents broadcast, and giving up a value r that might have been more beneficial than $V(r)$). The decision of which convention to use should be made based on solving for both cases and choosing the one yielding the maximum $V(r)$. One inherent advantage of broadcasting if below r is that it enables all agents to transmit in parallel, so that even if a collision occurs, a valid decision may be made since it is sufficient to know that at least one agent broadcasts (hence the exploration should be resumed).

4.2 Sequential Exploration

If the system is latency tolerant, a far more efficient (cost-wise) strategy can be derived, taking advantage of sequential exploration. In this case, instead of having all agents sense and broadcast the quality of a channel simultaneously, each agent dynamically decides, in its turn, whether to sense and broadcast-based on broadcasts received from other agents related to this channel. The idea is that if one of the agents experiences poor quality, then it is unnecessary to have all agents spend resources sensing that channel. Specifically, we define our strategy as a set of thresholds $S = \{r_1, r_2, .., r_k\}$. Agents sense and broadcast their quality for a specific channel sequentially, according to a predefined list. Since the qualities sensed by the different agents derive from the same distribution, there is no importance to the order according to which a channel is sensed by the different agents and any arbitrary pre-defined order is appropriate as long as all agents obey it. The sensing of a channel continues as long as the the minimum between the quality values reported by the i agents that have already sensed the channel is greater than a threshold r_i $(1 \leq i < k)$. Otherwise, if the minimum obtained after the $i - th$ $(1 \leq i < k)$ sensing is below r_i, the agents will switch to the next channel. If all k agents have sensed the channel, then a threshold r_k will be used to determine whether to: (a) terminate the exploration (i.e., choosing to use the current channel for the session), if the min-

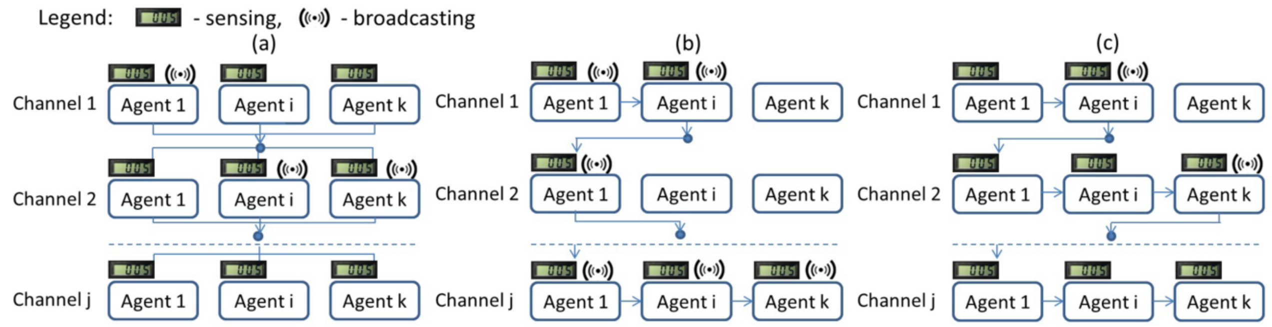

Figure 1. (a) "parallel+" - all agents perform sensing in parallel, and only those with quality below r broadcast. The process terminates when none of the agents broadcast. (b) "sequential" - each agent in its turn senses and broadcasts. The transition to the next channel is performed if a broadcast reveals a value lesser than r_i and the exploration terminates if none of the agents broadcast a value lesser than r_k. (c) "sequential+" - each agent i senses in its turn, however only broadcasts if below r_i. The transition to the next channel is performed if the $i-th$ agent broadcast and exploration terminates if no agent broadcasts in its turn.

imum found so far is above r_k; or otherwise (b) resume the exploration, switching to the next channel and using the same procedure. This process is illustrated in Figure 1(b).

We use $V(S)$ to denote the expected benefit of using strategy $S = \{r_1, r_2, .., r_k\}$. The thresholds in the expected-benefit maximizing strategy satisfy: $r_1 > r_2 > .. > r_{k-1} > r_k = V(S)$. Intuitively, this is explained by the fact that as the sensing process for a specific channel approaches the end of the chain, the probability for a further decrease in the minimum found so far becomes smaller, hence r_i decreases as i increases. Formally, this is proven by showing that any sequence $S' = \{r'_1, r'_2, .., r'_k\}$ where $r'_i < r'_{i+1}$ is dominated by a modification of that sequence in which $r'_i = r'_{i+1}$. The equality $r_k = V(S)$ is the essence of the threshold concept — if the minimum quality of the channel, based on all k sensing operations, is lower than the expected benefit of resuming exploration over the next channel ($V(S)$), then the exploration process should be resumed. Otherwise, the exploration process should be terminated and the current channel should be selected. The formal proof of this latter equality is omitted since it is a specific case of the proof given for Theorem 2 below.

Here again, the process can be further improved either by broadcasting above or below the threshold used (either using a binary signal or actually broadcasting the value sensed), and having the agents follow a convention according to which a broadcast is made by the ith agent only if its sensed quality is below a threshold r_i. The exploration of the same channel will be resumed as long as none of the agents have broadcast during their sensing turn. Once an agent broadcasts, the agents will switch to the next channel. The process is illustrated in Figure 1(c).

The expected benefit of using the improved sequential exploration strategy S (denoted "sequential+" onwards), $V(S)$, can be calculated recursively. We represent the state of the system by the pair (w, l), where l is the number of agents that have already sensed the specific channel and broadcast their observed quality, and w is the lowest quality sensed so far for that channel. We use $EB(w, l)$ to denote the expected benefit, onwards, from being in state (w, l) and using strategy S. The value of $EB(w, l)$ is in fact the expected minimum quality of the channel in which the exploration will terminate, minus the expected sum of costs accumulated along the process. Formally, the value of $EB(w, l)$ is given by:

$$EB(w,l) = -c_e + F(r_l)(V(S) - c_b) \quad (5)$$
$$+ \int_{y=r_l}^{w} EB(y, l+1)f(y)dy + (1-F(w))EB(w, l+1)$$

where $EB(w, k+1) = w$. Therefore, $V(S) = EB(\infty, 1)$.

Theorem 2. *The expected-benefit maximizing "sequential+" exploration strategy is a set $S = \{r_1, r_2, .., r_k\}$, which is the solution to the set of k equations:*

$$V(S) - c_b = EB(r_l, l+1), \qquad \text{for } l = 1, .., k. \quad (6)$$

Proof. Equating the first derivative of (5) to zero and applying some mathematical manipulation (in particular notice that $dV(S)/dr_l = 0$ for the expected-benefit maximizing r_l) obtains (6). □

Intuitively, (6) can be explained as follows. Each threshold r_l is in fact the quality value for which the system is indifferent between switching to exploring the next channel (which requires a broadcast and thus incurs a cost c_b), yielding expected benefit $V(S)$, and sticking with the current channel, yielding an expected benefit $EB(r_l, l+1)$ (or obtaining the value r_k if $l = k$).

Furthermore, substituting $EB(r_l, l+1)$ according to (5) in (6) obtains:

$$c_e = \int_{y=r_l}^{r_{l-1}} (EB(min(y, r_{l-1}), l+1) - (V(S) - c_b))f(y)dy \quad (7)$$

The value of the expected-benefit maximizing r_l can thus be interpreted as the quality value for which the cost of an additional exploration of the current channel equals the marginal benefit from executing such an exploration (rather than switching to the next channel).

The computational complexity of solving the set of k equations according to Theorem 2 highly depends on the distribution function $f(y)$. For some distribution functions (e.g., uniform or multi-rectangular) the integrals can be directly calculated, resulting in a set of linear equations that can be solved with a complexity of $o(k^3)$ (e.g., using Gaussian Elimination). For other distribution functions, numerical methods should be used.

The expected number of channels that will be explored when using the sequential exploration is $\eta_{channels} = \frac{1}{\prod_{i=1}^{k}(1-F(r_i))}$. This is equivalent to a Bernoulli sampling process with a success probability of having all values sensed be greater than the threshold used. The expected number of sensings for each channel that is not eventually selected is given by: $\eta_{sensings} = \frac{\sum_{i=1}^{k} iP(i)}{1-\prod_{i=1}^{k}(1-F(r_i))}$, where $P(i) = F(r_i)\prod_{l=1}^{i-1}(1 - F(r_l))$ denotes the probability that the agents switch to the next channel after the $i-th$ sensing. The overall expected latency is thus $(\eta_{channels} - 1)\eta_{sensings} + k$ (as k sensings need to be made for the last channel explored).

Before concluding this section we note that the regular sequential exploration discussed at the beginning of this section (i.e., when the agents always broadcast their findings) is a specific case solved by replacing (5) with: $EB(w, l) = -c_e - c_b + F(r_l)V(S) + \int_{y=r_l}^{w} EB(y, l+$

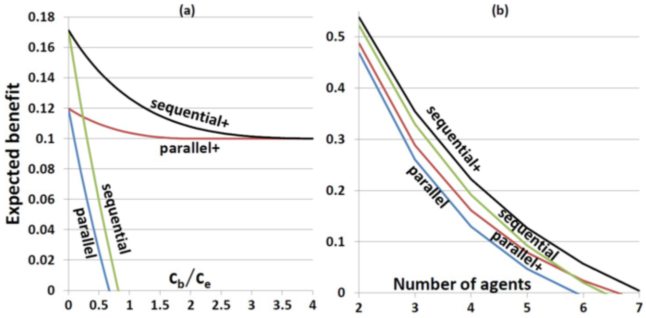

Figure 2. (a) Expected benefit of 3-agents' exploration as a function of the ratio c_b/c_e (keeping $c_e = 0.05$). (b) Expected benefit of k agents exploration as a function of k (using $c_b = 0.005$ and $c_e = 0.02$).

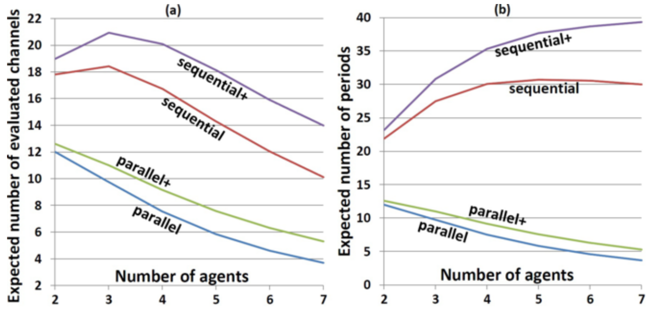

Figure 3. The expected number of channels explored (left) and the expected exploration latency (right) as a function of the number of agents k. The costs used are: $c_e = c_b = 0.002$.

$1)f(y)dy+(1-F(w))EB(w,l+1)$ and Theorem 2 holds in this case with the following modification of (6): $V(S) = EB(r_l, l+1)$.

Finally, we note that as an alternative to broadcasting only when the sensed quality is below r_l we can broadcast only if it is above r_l. In this case, the term $F(r_l)(V(S) - c_b)$ in (5) changes to $F(r_l)V(S) - (1 - F(r_l))c_b$ and (6) in Theorem 2 changes to $V(S) = EB(r_l, l+1) - c_b$. The method can thus be improved by individually setting a different convention for each agent's broadcast, indicating whether a broadcast or its absence indicate a value above or below a threshold (i.e., the agents with thresholds for which $F(r_l) < 0.5$ will broadcast only if receiving a value below r_l and vice versa). Seemingly, the solution in this case will require solving for all possible combinations of individual conventions and choosing the one associated with the maximum expected benefit. Nevertheless, since the thresholds necessarily decrease along the sequence, the transition in conventions can happen only once (and only from broadcasting if above threshold to broadcasting if below threshold). Therefore the number of combinations that needs to be solved for is linear in the number of agents.

5 Numerical Illustration

In order to illustrate the performance achieved with the different methods, we use a tractable synthetic setting that simplifies calculations, yet demonstrates the effectiveness and the cost latency tradeoffs encapsulated in the parallel and sequential exploration strategies. The setting uses a uniform distribution function defined over the interval $(0, 1)$.

We first demonstrate the effect of an increase in the number of agents that need to coordinate their exploration and the increase in the ratio between c_b and c_e over the expected benefit (see Figure 2). As expected, the "sequential+" method dominates all other methods, as far as expected benefit is concerned, and both the increase in exploration costs and in the number of agents that need to be coordinated result in a decrease in the expected benefit (in all methods). As the broadcast becomes the dominant among the two cost factors, the improvement achieved by selectively performing broadcast becomes more significant. The correlation between the expected benefit and the number of agents is explained by the fact that as the number of agents increases, more exploration needs to take place and the expected minimum of the observed quality of each channel decreases.

Figure 2(a) illustrates the overhead incurred by the costly coordination. The case of $c_b = 0$ is in fact the equivalence of free coordination, hence in this case there is no benefit in saving broadcasts (reflected by the similar result in both sequential-based methods and both parallel-based methods). As the value of c_b increases, the decrease in expected benefit fully measures the "cost of coordination" in each of the models.

Figure 3 illustrates the effect of changes in the number of agents exploring cooperatively over the expected number of channels the agents may request to explore and the expected resulting latency when using the different exploration methods. As expected, in the parallel exploration models both the number of channels evaluated and the latency decrease as the number of agents that need to coordinate their exploration increases. The reason for the decrease in both cases is the resulting decrease in the thresholds used by the agents, as reflected in (2) and (4) (indirectly due to the decrease in the expected minimum value attributed to a channel, $f_{min}(y)$, that results from the increase in the number of agents). A different behavior is observed in the sequential-based models. Here, when the number of agents increases, both the expected number of evaluated channels and the expected latency may increase up to a certain point and then decrease. This non-intuitive behavior is explained by the conflicting effects the increase in the number of agents has on the thresholds used when using sequential-based exploration: the increase in the number of agents results in a decrease in the thresholds used as discussed above. This, however, results in a decrease in the expected number of channels explored (and consequently in the overall latency), as the agents are likely to terminate their exploration even when encountering relatively small minimum values. On the other hand, since the number of agents increases, the chances of all agents sensing a quality above the thresholds used decreases also (despite the decrease in the thresholds used) and therefore more channels need to be explored. Furthermore, the fact that the lower thresholds are used suggests that as part of any channel's exploration more agents will perform sensing, on average, before a decision is made to switch to the next channel. This latter argument holds for the overall latency rather than the expected number of channels explored, hence explaining the fact that the decrease in the expected overall latency is observed for a greater number of agents (in comparison to the case with the expected number of channels).

Finally, Figure 4 depicts the $99.9th$ percentile of the number of channels that will be requested by the agents (i.e., the maximum number of channels that will need to be evaluated in 99.9% of the cases) as a function of the exploration and broadcast costs. The area below each curve represents settings in which in 0.1% of the cases the agents will request to explore more channels than its associated number states. From the figure we observe that for most reasonable values of c_e and c_b, the number of channels that will need to be evaluated is relatively moderate (e.g., in this example the probability of actually needing to explore more than 200 channels is practically insignificant). The importance of this observation is twofold: first, it justifies the use of a stationary set of thresholds as the probability that the number of channels that the agents will want to evaluate will exceed the number of available channels is, in most settings, negligible. Second, it reassures that the four methods are applicable

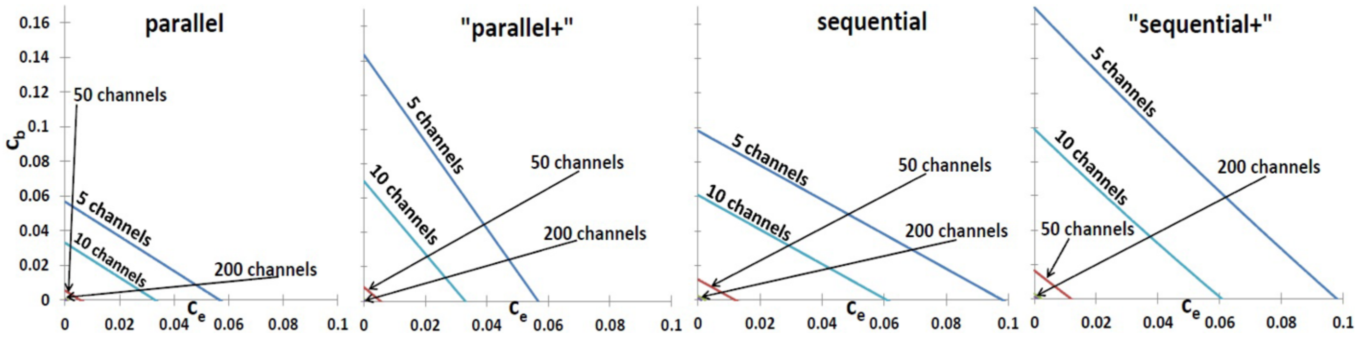

Figure 4. The maximum number of channels that the agent will request to evaluate in 99.9% of the cases (the $99.9th$ percentile), in the different exploration methods. Each curve represents the set of (c_b, c_e) values for which the $99.9th$ percentile number of channels that will be requested to evaluate equals the value associated with that curve. The number of agents used is $k = 3$.

latency-wise.

6 Conclusion

The need to coordinate multi-agent exploration in a distributed manner arises whenever a group of agents engage in costly exploration and individual goals are governed by a common function of all findings. The offload cognitive channel exploration coordinated among secondary users sharing a common control channel with scarce resources is just one out of many examples of this kind.

From Section 5 we learn that costly coordination can substantially worsen the performance (in terms of expected benefit) in distributed cooperative multi-agent exploration applications. The performance can be enhanced by reducing the amount of information that needs to be broadcasted (if at all); instead of broadcasting the sensed individual quality, the agents broadcast only upon finding a quality that is above or below pre-defined thresholds and interpret broadcasts and their absence according to a predetermined protocol. All the strategies proposed in this paper can be calculated offline with a polynomial complexity and all the online processing involves merely threshold-based decisions. The different exploration schemes differ mainly in the tradeoff they offer between expected benefit and latency. Future work includes introducing minimum overall quality constraint and additional latency effective models. For example, any of the methods can be tuned to guarantee better latency by decreasing the thresholds used. Another promising direction involves sensing several channels prior to broadcasting, considering the tradeoff between the additional costs of potentially unnecessary sensing and the potential savings on broadcasts.

Acknowledgments

This work was partially supported by ISF/BSF grants 1401/09 and 2008-404, the Israeli Ministry of Industry and Trade under project RESCUE. The authors would like to thank Ron Rotstein, Kobi J. Scheim, and Nadav Lavi from General Motors, Advanced Technical Center - Israel, for fruitful discussions and insightful review.

REFERENCES

[1] I. Akyildiz, B. Lo, and R. Balakrishnan, 'Cooperative spectrum sensing in cognitive radio networks: A survey', *Physical Comm.*, **4**, 40–62, (2011).
[2] M. Chhabra, S. Das, and D. Sarne, 'Expert-mediated search', in *Proc. AAMAS*, pp. 415–422, (2011).
[3] C. Chou, S. Sai, H. Kim, and K. Shin, 'What and How Much to Gain by Spectrum Agility?', *IEEE J-SAC*, **25**, 576–588, (2007).
[4] M. D. Felice, K.R. Chowdhury, and L. Bononi, 'Analyzing the potential of cooperative cognitive radio technology on inter-vehicle communication', in *Proc. Wireless Days*, pp. 1–6, (2010).
[5] A. Frieder, R. Lin, and S. Kraus, 'Agent-human coordination with communication costs under uncertainty', in *Proc. AAAI-12*, to appear, (2012).
[6] M. Gandetto and C. Regazzoni, 'Spectrum Sensing: A Distributed Approach for Cognitive Terminals', *IEEE J-SAC*, **25**, 546–557, (2007).
[7] J. Gatti, 'Multi-commodity consumer search', *JET*, **86**, 219–244, (1999).
[8] A. Grosfeld-Nir, D. Sarne, and I. Spiegler, 'Modeling the search for the least costly opportunity', *EJOR*, **197**, 667–674, (2009).
[9] B. J. Grosz, L. Hunsberger, and S. Kraus, 'Planning and acting together', *AI Magazine*, **20**, 23–34, (1999).
[10] N. Hazon, Y. Aumann, and S. Kraus, 'Collaborative multi agent physical search with probabilistic knowledge', in *Proc. IJCAI*, pp. 167–174, (2009).
[11] S. Kraus, R. Lin, and Y. Shavitt, 'On self-interested agents in vehicular networks with car-to-car gossiping', *IEEE Trans. Vehic. Tech.*, **57**, 3319–3332, (2008).
[12] J. McMillan and M. Rothschild, 'Search', in *Proc. Handbook of Game Theory with Economic Applications*, volume 2, pp. 905–927, (1994).
[13] P. Michiardi and R. Molva, 'Analysis of coalition formation and cooperation strategies in mobile ad hoc networks', *Ad Hoc Net.*, **3**, 193–219, (2005).
[14] P. Morgan, 'Search and optimal sample sizes', *Review of Economic Studies*, **50**, 659–675, (1983).
[15] P. Morgan and R. Manning, 'Optimal search', *Econometrica*, **53**, 923–944, (1985).
[16] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, 'Optimal multiband joint detection for spectrum sensing in cognitive radio networks', *IEEE Trans. on Signal Processing*, **57**, 1128–1140, (2009).
[17] I. Rochlin, D. Sarne, and G. Zussman, 'Sequential multilateral search for a common goal', in *Proc. IAT*, pp. 349–356, (2011).
[18] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, 'Coalitional games for distributed collaborative spectrum sensing in cognitive radio networks.', *CoRR*, 2114–2122, (2009).
[19] T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohme, 'Coalition structure generation with worst case guarantees', *AIJ*, **111**, 209–238, (1999).
[20] D. Sarne, E. Manisterski, and S. Kraus, 'Multi-goal economic search using dynamic search structures', *JAAMAS*, **21**, 204–236, (2010).
[21] S. Xie, Y. Liu, Y. Zhang, and R. Yu, 'A parallel cooperative spectrum sensing in cognitive radio networks', *IEEE Trans. Vehic. Tech.*, **59**, 4079–4092, (2010).
[22] Xiao Y. and P. Ho, 'A novel sensing coordination framework for cr-vanets', *IEEE Trans. Vehic. Tech.*, **59**, 1936 –1948, (2010).
[23] J. Yamamoto and K. Sycara, 'A stable and efficient buyer coalition formation scheme for e-marketplaces', in *Proc. Agents'01*, pp. 576–583, (2001).